

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: June 5, 2006, 12:31:47 ; Search time 97.4247 Seconds
(without alignments)
65.702 Million cell updates/sec

Title: US-10-645-659A-10
Perfect score: 86
Sequence: 1 TWHYYLNGRTATR 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : A_Geneseq_8:*

1: Geneseq1990s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*
9: Geneseq2005s:.*
10: Geneseq2006s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	100.0	14	8	ADR88216 Human hep
2	86	100.0	14	8	ADT78183 Functiona
3	86	100.0	14	9	AEA42432 Human hep
4	86	100.0	15	9	ADU71059 Human hep
5	86	100.0	15	9	ADU71060 Human hep
6	86	100.0	15	9	ADU727047 Heparanas
7	86	100.0	386	8	ADR88207 Human mat
8	86	100.0	386	8	ADT78174 45KDa sub
9	86	100.0	386	9	ADY27057 Heparanas
10	86	100.0	386	9	ADZ18995 Human hep
11	86	100.0	386	9	AEA42423 Human mat
12	86	100.0	460	9	ADY27061 Heparanas
13	86	100.0	486	9	AE887589 Human hep
14	86	100.0	492	9	ADZ18996 Hep106 co
15	86	100.0	493	9	AE887562 Human hep
16	86	100.0	495	9	ADZ18999 Human hep
17	86	100.0	497	9	AE887587 Human hep
18	86	100.0	501	9	ADZ19000 HepGS3 co
19	86	100.0	507	9	ADZ19005 HepGS6 co
20	86	100.0	508	9	ADY27058 Human ina
21	86	100.0	526	9	ADZ19006 HepHyalur
22	86	100.0	527	5	ABB07815 Chicken s
23	86	100.0	527	7	ABW02018 Chimeric

24	86	100.0	527	8	ADO63825	ADO63825 Chimeric
25	86	100.0	527	8	ADO63827	ADO63827 Chimeric
26	86	100.0	527	8	ADO63826	ADO63826 Chimeric
27	86	100.0	527	9	ADZ19004	ADZ19004 HepGS4 co
28	86	100.0	532	2	AA171083	AA171083 Seq ID No
29	86	100.0	543	2	AA17082	AA17082 A human h
30	86	100.0	543	2	AA17082	AA17082 Human hep
31	86	100.0	543	3	AA17082	AA17082 Human hep
32	86	100.0	543	3	AA17082	AA17082 Human hep
33	86	100.0	543	3	AA17082	AA17082 Human hep
34	86	100.0	543	4	AA17082	AA17082 Human hep
35	86	100.0	543	4	AA17082	AA17082 Human hep
36	86	100.0	543	4	AA17082	AA17082 Human hep
37	86	100.0	543	5	AA17082	AA17082 Human hep
38	86	100.0	543	7	AA17082	AA17082 Human hep
39	86	100.0	543	7	AA17082	AA17082 Human hep
40	86	100.0	543	8	AA17082	AA17082 Human hep
41	86	100.0	543	8	AA17082	AA17082 Human hep
42	86	100.0	543	8	AA17082	AA17082 Human hep
43	86	100.0	543	8	AA17082	AA17082 Human hep
44	86	100.0	543	8	AA17082	AA17082 Human hep
45	86	100.0	543	8	AA17082	AA17082 Human hep
46	86	100.0	543	8	AA17082	AA17082 Human hep
47	86	100.0	543	8	AA17082	AA17082 Human hep
48	86	100.0	543	8	AA17082	AA17082 Human hep
49	86	100.0	543	8	AA17082	AA17082 Human hep
50	86	100.0	543	8	AA17082	AA17082 Human hep
51	86	100.0	543	8	AA17082	AA17082 Human hep
52	86	100.0	543	8	AA17082	AA17082 Human hep
53	86	100.0	543	8	AA17082	AA17082 Human hep
54	86	100.0	543	8	AA17082	AA17082 Human hep
55	86	100.0	543	9	AA17082	AA17082 Human hep
56	86	100.0	543	9	AA17082	AA17082 Human hep
57	86	100.0	543	9	AA17082	AA17082 Human hep
58	86	100.0	543	9	AA17082	AA17082 Human hep
59	86	100.0	543	10	AA17082	AA17082 Human hep
60	86	100.0	545	6	AA17082	AA17082 Human hep
61	86	100.0	545	8	AA17082	AA17082 Human hep
62	86	100.0	545	8	AA17082	AA17082 Human hep
63	86	100.0	556	9	AA17082	AA17082 Human hep
64	86	100.0	570	9	AA17082	AA17082 Human hep
65	86	100.0	588	2	AA17082	AA17082 Human hep
66	86	100.0	592	2	AA17082	AA17082 Human hep
67	86	100.0	592	3	AA17082	AA17082 Human hep
68	86	100.0	592	7	AA17082	AA17082 Human hep
69	86	100.0	592	8	AA17082	AA17082 Human hep
70	86	100.0	592	8	AA17082	AA17082 Human hep
71	86	100.0	592	9	AA17082	AA17082 Human hep
72	86	100.0	592	9	AA17082	AA17082 Human hep
73	86	100.0	592	9	AA17082	AA17082 Human hep
74	86	100.0	592	9	AA17082	AA17082 Human hep
75	86	100.0	592	9	AA17082	AA17082 Human hep
76	86	100.0	592	9	AA17082	AA17082 Human hep
77	86	100.0	592	9	AA17082	AA17082 Human hep
78	86	100.0	592	9	AA17082	AA17082 Human hep
79	86	100.0	592	9	AA17082	AA17082 Human hep
80	86	100.0	592	9	AA17082	AA17082 Human hep
81	86	100.0	592	9	AA17082	AA17082 Human hep
82	86	100.0	592	9	AA17082	AA17082 Human hep
83	86	100.0	592	9	AA17082	AA17082 Human hep
84	86	100.0	592	9	AA17082	AA17082 Human hep
85	86	100.0	592	9	AA17082	AA17082 Human hep
86	86	100.0	592	9	AA17082	AA17082 Human hep
87	86	100.0	592	9	AA17082	AA17082 Human hep
88	86	100.0	592	9	AA17082	AA17082 Human hep
89	86	100.0	592	9	AA17082	AA17082 Human hep
90	86	100.0	592	9	AA17082	AA17082 Human hep
91	86	100.0	592	9	AA17082	AA17082 Human hep
92	86	100.0	592	9	AA17082	AA17082 Human hep
93	86	100.0	592	9	AA17082	AA17082 Human hep
94	86	100.0	592	9	AA17082	AA17082 Human hep
95	86	100.0	592	9	AA17082	AA17082 Human hep
96	86	100.0	592	9	AA17082	AA17082 Human hep

97	74	86.0	545	9	ADY27034	Ady27034 Bovine he	170	43	50.0	981	5	ABB78727	Abb78727 Human cal
98	72	83.7	15	9	ADU70969	Adu70969 Human hep	171	43	50.0	981	7	ADe40182	AdE40182 Human NOV
99	70	81.4	15	9	ADU70927	Adu70927 Human hep	172	43	50.0	981	9	ADY70315	AdY70315 Human bet
100	67	77.9	15	9	ADU71058	Adu71058 Human hep	173	43	50.0	981	9	ADY70649	AdY70649 Human BAC
101	62	72.1	15	9	ADU70797	Adu70797 Human hep	174	43	50.0	1004	4	ABb65405	AbB65405 Drosophil
102	62	72.1	15	9	ADU70662	Adu70662 Human hep	175	43	50.0	1027	6	ABP75726	AbP75726 Human sec
103	62	72.1	15	9	ADU71272	Adu71272 Human hep	176	43	50.0	1177	8	AAB96721	AaB96721 Putative
104	62	72.1	15	9	ADU71210	Adu71210 Human hep	177	43	50.0	1177	8	ADs43128	AdS43128 Bacterial
105	56	65.1	9	9	ADU70700	Adu70700 Human hep	178	43	50.0	1200	8	ADN20018	AdN20018 Bacterial
106	56	65.1	9	9	ADU70638	Adu70638 Human hep	179	42.5	49.4	1035	7	ADc00940	AdC00940 Enterohae
107	56	65.1	15	9	ADU70968	Adu70968 Human hep	180	42.5	49.4	1035	9	AEd81975	AeD81975 Hyperimmu
108	56	65.1	50	5	AAm50385	Aam50385 Mouse hep	181	42	48.8	56	4	AAU63539	AaU63539 Propionib
109	54	62.8	15	9	ADU71211	Adu71211 Human hep	182	42	48.8	56	6	ABM60058	ABm60058 Propionib
110	52	60.5	9	9	ADU70639	Adu70639 Human hep	183	42	48.8	218	6	ABU21887	ABu21887 Protein e
111	51	59.3	1179	8	ADs41964	AdS41964 Bacterial	184	42	48.8	223	4	AAU64792	AAu64792 Propionib
112	51	59.3	1179	8	ADN18677	Adn18677 Bacterial	185	42	48.8	223	6	ABM61311	ABm61311 Propionib
113	50	58.1	9	9	ADU70390	Adu70390 Human hep	186	42	48.8	297	4	ABBS2977	ABbS2977 Escherich
114	50	58.1	15	9	ADU70967	Adu70967 Human hep	187	42	48.8	322	8	ADX67425	ADx67425 Plant ful
115	49	57.0	9	9	ADU70516	Adu70516 Human hep	188	42	48.8	337	6	ABUS2328	ABuS2328 Human GPC
116	47	54.7	9	9	ADU70342	Adu70342 Human hep	189	42	48.8	337	8	ADL23991	ADl23991 Human NOV
117	47	54.7	15	9	ADU71061	Adu71061 Human hep	190	42	48.8	337	8	ABO84700	ABo84700 Mouse can
118	46	53.5	15	9	ADU71209	Adu71209 Human hep	191	42	48.8	340	2	AAr65244	AaR65244 Human SF3
119	46	53.5	179	6	ABR40721	Abr40721 Momordica	192	42	48.8	340	7	ADb75541	ADb75541 Prostata
120	46	53.5	179	7	ADC23627	Adc23627 Polypepti	193	42	48.8	340	7	ADD14165	ADd14165 Human src
121	46	53.5	245	8	ADN24954	Adn24954 Bacterial	194	42	48.8	340	8	ABO84704	ABo84704 Human can
122	46	53.5	245	8	ADN22195	Adn22195 Bacterial	195	42	48.8	340	8	ABO84703	ABo84703 Human can
123	46	53.5	314	2	AAr75200	Aar75200 Rat P-F4M	196	42	48.8	340	8	ABO84702	ABo84702 Human can
124	46	53.5	350	2	AAr75198	Aar75198 Rat Gal-b	197	42	48.8	340	8	ABO84701	ABo84701 Human can
125	46	53.5	350	7	ADe59126	AdE59126 Rat Prote	198	42	48.8	343	2	AAr41670	AaR41670 Porcine s
126	46	53.5	350	7	ADe59130	AdE59130 Rat Prote	199	42	48.8	343	2	AAr65240	AaR65240 Porcine s
127	46	53.5	350	7	ADe64099	AdE64099 Human Pro	200	42	48.8	343	9	AED08895	AeD08895 Amino aci
128	46	53.5	350	7	ADe64097	AdE64097 Rat Prote	201	42	48.8	358	6	ABM64563	ABm64563 Propionib
129	46	53.5	350	7	ADd45580	AdD45580 Rat Prote	202	42	48.8	358	8	ADN60340	ADn60340 B. lichen
130	46	53.5	350	7	ABM85655	ABm85655 Human pro	203	42	48.8	375	8	ADN17709	ADn17709 Bacterial
131	46	53.5	351	7	ADe59132	AdE59132 Human pro	204	42	48.8	399	4	ABG18000	ABg18000 Novel hum
132	46	53.5	351	7	ADd45582	AdD45582 Human pro	205	42	48.8	554	7	ADe56932	ADe56932 Rat Prote
133	46	53.5	351	7	ADe59128	AdE59128 Human pro	206	42	48.8	554	7	ADe56928	ADe56928 Rat Prote
134	46	53.5	379	7	ABM85654	ABm85654 Mouse pro	207	42	48.8	574	2	AAr66033	AaR66033 Human ARD
135	44	51.2	2835	5	ABb98574	Abb98574 Dextran s	208	42	48.8	574	2	AAr66034	AaR66034 Rat ARD 1
136	44	51.2	2835	6	ABR55594	Abr55594 Amino aci	209	42	48.8	574	7	ADe56930	ADe56930 Human Pro
137	44	51.2	3972	3	AAb23749	Aab23749 S. avermi	210	42	48.8	574	7	ADe56934	ADe56934 Human Pro
138	44	51.2	3972	4	AAg65264	Aag65264 Streptomy	211	42	48.8	664	4	AAb93120	ABa93120 Human pro
139	44	51.2	3972	4	AAg65268	Aag65268 Streptomy	212	42	48.8	664	5	AAU10928	AAu10928 Protein o
140	44	51.2	5532	3	AAb23752	Aab23752 S. avermi	213	42	48.8	664	5	AAU77036	AAu77036 Human TRC
141	44	51.2	5532	4	AAg65267	Aag65267 Streptomy	214	42	48.8	664	6	ABJ37044	ABj37044 Human bre
142	43	50.0	181	4	AAb59819	Aab59819 Tute prot	215	42	48.8	664	7	ADb80932	ADb80932 RING-SH c
143	43	50.0	331	5	AAm50383	Aam50383 Human hep	216	42	48.8	664	8	ADR89544	ADr89544 Apoptosis
144	43	50.0	437	9	ABM90536	ABm90536 M. xanthu	217	42	48.8	664	8	ADx06417	ADx06417 Cyclin-de
145	43	50.0	439	4	AAU07423	Aau07423 Human hep	218	42	48.8	668	8	ADR89542	ADr89542 Apoptosis
146	43	50.0	470	5	AAE18328	Aae18328 Human hep	219	42	48.8	691	9	AED08889	AeD08889 Amino aci
147	43	50.0	480	4	AAy97634	Aay97634 Human hep	220	42	48.8	841	9	AED08890	AeD08890 Amino aci
148	43	50.0	480	4	AAU07418	Aau07418 Novel hum	221	42	48.8	956	5	ABb78729	ABb78729 Human cal
149	43	50.0	480	4	AAb85217	Aab85217 Heparanas	222	42	48.8	956	9	ADY70317	ADy70317 Human bet
150	43	50.0	492	4	AAb84664	Aab84664 Amino aci	223	42	48.8	968	7	ADc65993	ADc65993 Human XB3
151	43	50.0	528	5	AAE18327	Aae18327 Human hep	224	42	48.8	968	9	ADz25519	ADz25519 Cytokine
152	43	50.0	534	5	ABP69310	ABp69310 Heparanas	225	42	48.8	968	9	AEa15828	AeA15828 Human alc
153	43	50.0	534	5	ABP69310	ABp69310 Human pol	226	42	48.8	968	9	AEc21698	AeC21698 Alkadiene
154	43	50.0	534	5	AAy97633	Aay97633 Human pre	227	42	48.8	1022	9	ADz25508	ADz25508 Cytokine
155	43	50.0	538	4	AAy97633	Aay97633 Human hep	228	42	48.8	1144	9	AED09961	AeD09961 Pathogeni
156	43	50.0	581	8	ADu25303	ADu25303 Bacillus	229	42	48.8	1175	9	AED82435	AeD82435 Hyperimmu
157	43	50.0	582	5	AAE18326	Aae18326 Human hep	230	41	47.7	120	6	ABU31307	ABu31307 Protein e
158	43	50.0	592	4	AAy97632	Aay97632 Human hep	231	41	47.7	133	4	AAU54261	AaU54261 Propionib
159	43	50.0	592	4	AAU07424	Aau07424 Human hep	232	41	47.7	133	6	ABM50780	ABm50780 Propionib
160	43	50.0	592	4	AAU07424	Aau07424 Human hep	233	41	47.7	179	4	AAU18319	AaU18319 Human end
161	43	50.0	592	4	AAb85215	ABa85215 Heparanas	234	41	47.7	202	7	ABM89756	ABm89756 Rice abio
162	43	50.0	608	6	ABU20989	ABu20989 Protein e	235	41	47.7	226	8	ADN99749	ADn99749 Novel hum
163	43	50.0	953	7	ADe28107	Ade28107 Human NTR	236	41	47.7	240	8	ADx87351	ADx87351 Plant ful
164	43	50.0	962	7	ADe28106	Ade28106 Human NTR	237	41	47.7	331	8	ADT58484	ADt58484 Plant pol
165	43	50.0	971	7	ADc65991	ADc65991 Human XB3	238	41	47.7	416	6	ABU32312	ABu32312 Protein e
166	43	50.0	971	9	AEa15827	AeA15827 Human BAC	239	41	47.7	428	2	AAr44403	AaR44403 Canine co
167	43	50.0	971	9	AEa15827	AeA15827 Human alc	240	41	47.7	435	7	ABO67223	ABo67223 Klebsiell
168	43	50.0	971	9	AEb06096	Aeb06096 Amino aci	241	41	47.7	479	6	ABP79127	ABp79127 N. gonorr
169	43	50.0	971	9	AEc21696	Aec21696 Alkadiene	242	41	47.7	490	5	ABP65335	ABp65335 Bifidobac

243	41	47.7	507	8	ADK16983	Nanoarcha	316	40	46.5	572	5	ABB47327	Listeria
244	41	47.7	511	4	AAU38104	Streptoco	317	40	46.5	646	4	ABG06358	Novel hum
245	41	47.7	511	9	ADY40315	iPGM Sco.	318	40	46.5	670	8	ADO61975	Transcrip
246	41	47.7	578	8	ADQ65875	Novel hum	319	40	46.5	720	8	ADN23654	Bacterioc
247	41	47.7	609	8	ADQ66688	Novel hum	320	40	46.5	1191	8	ADN46644	Thermoco
248	41	47.7	610	4	AAAB31664	Amino aci	321	40	46.5	2054	4	ABG23323	Novel hum
249	41	47.7	748	2	AAR24396	Prod. of	322	40	46.5	2273	4	ABG23304	Novel hum
250	41	47.7	748	2	AAR24398	Prod. of	323	40	46.5	5909	4	ABG233295	Novel hum
251	41	47.7	748	2	AAR24513	Consensus	324	40	46.5	6619	4	ABG23329	Novel hum
252	41	47.7	749	2	AAR24513	Feline in	325	39	45.3	9	ADU70763	Human hep	
253	41	47.7	749	2	AAR24513	Feline in	326	39	45.3	15	9	ADU71057	Human hep
254	41	47.7	819	8	ADP99142	Human tra	327	39	45.3	38	4	AAU17157	Peptide #
255	41	47.7	908	3	AAAB42511	Human ORF	328	39	45.3	38	4	ABG36158	Peptide #
256	41	47.7	1101	2	AAR44401	Canine co	329	39	45.3	38	4	AAU29648	Peptide #
257	41	47.7	1225	8	ADU08938	Coronavir	330	39	45.3	38	4	ABG30966	Peptide #
258	41	47.7	1307	6	AAO31015	Human tra	331	39	45.3	38	4	ABG21544	Protein #
259	41	47.7	1353	6	AAE29913	Human tra	332	39	45.3	38	4	AAU69327	Human bon
260	41	47.7	1359	8	ADG75981	Human ATP	333	39	45.3	38	4	ABG51002	Human liv
261	41	47.7	1391	8	ADP99172	Human tra	334	39	45.3	38	4	AAU04851	Peptide #
262	41	47.7	1443	2	AAR27818	CCV-6 spi	335	39	45.3	38	5	ABG38943	Human pep
263	41	47.7	1447	8	ABY00038	Transmiss	336	39	45.3	64	7	ABO65501	Klebsiell
264	41	47.7	1449	8	ADN37292	PTGV spik	337	39	45.3	85	4	ABG30007	Novel hum
265	41	47.7	1451	2	AAR27819	CCVinsavc	338	39	45.3	90	4	AAE10718	Chicken 4
266	41	47.7	1452	2	AAR24271	Canine co	339	39	45.3	90	4	AAE10717	Pig 4ST3G
267	41	47.7	1452	2	AAR44400	Canine co	340	39	45.3	90	4	AAE10713	Rat 4ST3G
268	41	47.7	1452	6	ABP98183	Amino aci	341	39	45.3	90	4	AAE10716	Mouse 4ST
269	41	47.7	1453	2	AAR27820	CCV-CS4 s	342	39	45.3	90	4	AAE10715	Human 4ST
270	41	47.7	1453	8	AAU31038	E2 protei	343	39	45.3	122	3	AAU41865	Human ORF
271	41	47.7	1453	8	ADU08932	Coronavir	344	39	45.3	133	4	AAU66016	Propionib
272	41	47.7	1454	2	AAR24511	Prod. of	345	39	45.3	133	6	ABM62535	Propionib
273	41	47.7	1454	2	AAR24397	Prod. of	346	39	45.3	133	6	ABM62535	Pseudomon
274	41	47.7	1454	2	AAR24395	Prod. of	347	39	45.3	149	7	ABO83324	Pseudomon
275	41	47.7	1454	2	AAR24468	Feline en	348	39	45.3	173	4	AAU92528	Human dig
276	41	47.7	1454	2	AAR24277	FECV/PIPV	349	39	45.3	173	4	AAU22565	Novel hum
277	41	47.7	1454	2	AAR24273	FIPV/FECV	350	39	45.3	173	7	ADB32405	Human nov
278	41	47.7	1454	2	AAR24274	FECV/PIPV	351	39	45.3	186	4	ABU53119	Intracell
279	41	47.7	1454	2	AAR24272	FECV/PIPV	352	39	45.3	190	8	ADR10089	Human pro
280	41	47.7	1454	2	AAR24267	Feline in	353	39	45.3	230	8	ADQ82666	Wild type
281	41	47.7	1454	2	AAR24270	Feline in	354	39	45.3	240	4	ABG65939	Drosophil
282	41	47.7	1454	2	AAR24275	FECV/PIPV	355	39	45.3	273	3	AAU27984	Arabidops
283	41	47.7	1454	2	AAR24278	FECV/PIPV	356	39	45.3	277	3	AAU27983	Arabidops
284	41	47.7	1454	2	AAR24264	Feline in	357	39	45.3	284	3	AAU07300	Arabidops
285	41	47.7	1483	1	AAU80474	Sequence	358	39	45.3	293	8	ADJ48984	Oil-associ
286	41	47.7	1594	1	AAU81183	Sequence	359	39	45.3	294	7	ABO43093	A. thalia
287	41	47.7	6239	3	AAU23750	S. averm	360	39	45.3	294	7	ADU31859	Plant (A.
288	41	47.7	6239	4	AAU65265	Streptomy	361	39	45.3	294	7	ADU55772	Thalecres
289	40	46.5	15	9	ADU70798	Human hep	362	39	45.3	294	8	ADU02279	Thalecres
290	40	46.5	27	4	AAU67973	Modified	363	39	45.3	295	3	AAU07299	Arabidops
291	40	46.5	47	6	ABP77982	N. gonorr	364	39	45.3	305	4	ABG13155	Novel hum
292	40	46.5	66	3	AAU57466	Arabidops	365	39	45.3	306	4	ABG08885	Novel hum
293	40	46.5	66	3	AAU59580	Arabidops	366	39	45.3	314	4	ABG13158	Novel hum
294	40	46.5	74	3	AAU73012	Neisseria	367	39	45.3	320	6	ABU37049	Protein e
295	40	46.5	81	6	ABP77980	N. gonorr	368	39	45.3	321	3	AAU27982	Arabidops
296	40	46.5	115	7	ADU08017	Bacterial	369	39	45.3	326	6	ABU25302	Protein e
297	40	46.5	144	6	ABU19391	Human int	370	39	45.3	365	4	AAU96750	Pucative
298	40	46.5	244	5	ABU90262	Human pol	371	39	45.3	389	8	ADT60416	Plant pol
299	40	46.5	338	6	ABM68047	Phototrab	372	39	45.3	389	9	ABE14376	Plant lip
300	40	46.5	349	3	AAU51208	Arabidops	373	39	45.3	426	8	ADL04592	M. catarr
301	40	46.5	353	3	AAU51207	Arabidops	374	39	45.3	451	5	ABG92483	Herbicida
302	40	46.5	374	8	ADN25772	Bacterial	375	39	45.3	452	5	ABG93908	Herbicida
303	40	46.5	374	8	ADN25772	Bacterial	376	39	45.3	478	4	ABG08884	Novel hum
304	40	46.5	375	8	ADN25772	Bacterial	377	39	45.3	478	4	ABG13156	Novel hum
305	40	46.5	389	6	ABM68309	Phototrab	378	39	45.3	529	2	AAU82586	Human pat
306	40	46.5	390	6	ABU22412	Protein e	379	39	45.3	551	6	ABU28836	Protein e
307	40	46.5	396	8	ADG66961	Plant ful	380	39	45.3	558	6	AAU33583	Acinetoba
308	40	46.5	397	3	AAU51206	Arabidops	381	39	45.3	572	6	AAU33683	Human str
309	40	46.5	400	6	ABU23209	Protein e	382	39	45.3	576	4	ABE14241	Drosophil
310	40	46.5	464	8	ADN36976	Plant ful	383	39	45.3	583	4	ABG63650	Drosophil
311	40	46.5	508	7	ADN03819	Human pro	384	39	45.3	690	4	ABG63232	Drosophil
312	40	46.5	508	9	AEC86749	Human cdn	385	39	45.3	809	9	ADN05059	Horse IL4
313	40	46.5	509	4	ABG23315	Novel hum	386	39	45.3	809	9	ADN05094	Horse IL4
314	40	46.5	534	9	ADW95051	Lycopersi	387	39	45.3	809	9	ADN05095	Horse IL4
315	40	46.5	534	9	ADX05253	Lycopersi	388	39	45.3	809	9	ADN05093	Horse IL4

389	39	45.3	809	9	ADX05046	Adx05046	Horse IL4	462	38	44.2	256	2	AA08564	Aay08564	B. subtil
390	39	45.3	822	4	ABG24362	Abg24362	Novel hum	463	38	44.2	260	8	ADX93633	Adx93633	Plant ful
391	39	45.3	822	4	ABG29163	Abg29163	Novel hum	464	38	44.2	275	6	ADA36243	Ada36243	Acinetoba
392	39	45.3	822	4	ABG08883	Abg08883	Novel hum	465	38	44.2	276	8	ADO80530	Ado80530	Burkholde
393	39	45.3	825	7	ABO68417	Ab068417	Pseudomon	466	38	44.2	285	4	AA095183	Aab95183	Human pro
394	39	45.3	861	4	ABG22537	Abg22537	Novel hum	467	38	44.2	285	7	ADC31422	Adc31422	Human nov
395	39	45.3	861	4	ABG18082	Abg18082	Novel hum	468	38	44.2	289	8	ADN73295	Adn73295	Thale cre
396	39	45.3	887	6	ABU45334	Abu45334	Protein e	469	38	44.2	289	8	ADN73295	Adn73295	Thale cre
397	39	45.3	889	6	ABU46941	Abu46941	Protein e	470	38	44.2	301	8	ADY09051	Ady09051	Plant ful
398	39	45.3	889	6	ABU47500	Abu47500	Protein e	471	38	44.2	301	8	ABU07872	Abu07872	Human zin
399	39	45.3	890	6	ABU15116	Abu15116	Protein e	472	38	44.2	313	4	AA090653	Aab90653	Human sec
400	39	45.3	959	8	ABO84640	Ab084640	Human can	473	38	44.2	326	4	AA090652	Aab90652	Human sec
401	39	45.3	960	8	ABO84643	Ab084643	Human can	474	38	44.2	353	8	ADS44665	Ads44665	Bacterial
402	39	45.3	971	7	ADD46684	Ad46684	Rat Prote	475	38	44.2	385	3	AA019189	Aab19189	Lipid met
403	39	45.3	986	8	ABO84645	Ab084645	Human can	476	38	44.2	385	9	ADV70586	Adv70586	Brevibact
404	39	45.3	1007	7	ADD01193	Ad01193	Human nuc	477	38	44.2	385	9	ADM15128	Adm15128	B. lactof
405	39	45.3	1013	8	ABO84644	Ab084644	Human can	478	38	44.2	395	8	ADS21844	Ads21844	Bacterial
406	39	45.3	1017	8	ABO84642	Ab084642	Human can	479	38	44.2	398	8	ADS41600	Ads41600	Bacterial
407	39	45.3	1018	8	ABO84641	Ab084641	Human can	480	38	44.2	402	4	ABB50185	Abb50185	Human tra
408	39	45.3	1113	6	ABB82999	Abb82999	Human KCN	481	38	44.2	402	8	ADJ64309	Adj64309	Cartilage
409	39	45.3	1146	3	AA092225	Aay92225	Human pat	482	38	44.2	402	9	AED26270	Aed26270	Novel hum
410	39	45.3	1154	6	ABO07240	Ab007240	Human p53	483	38	44.2	403	2	AAW72943	Aaw72943	Mycobacte
411	39	45.3	1154	6	ABB83000	Abb83000	Human KCN	484	38	44.2	403	2	AA021963	Aay21963	Amino aci
412	39	45.3	1154	7	ADD46686	Ad46686	Human Pro	485	38	44.2	404	2	AAW72942	Aaw72942	Mycobacte
413	39	45.3	1154	8	ADJ75670	Adj75670	Marker ge	486	38	44.2	404	2	AA021962	Aay21962	Amino aci
414	39	45.3	1171	8	ADJ76368	Adj76368	Marker ge	487	38	44.2	410	8	ADJ64307	Adj64307	Cartilage
415	39	45.3	1171	8	ABO84639	Ab084639	Mouse can	488	38	44.2	419	4	ABB66861	Abb66861	Drosophil
416	39	45.3	1172	8	ADO59149	Ado59149	Chicken S	489	38	44.2	424	4	ABB61372	Abb61372	Drosophil
417	39	45.3	1174	7	ABM79021	Abm79021	BK channe	490	38	44.2	483	5	ABB76634	Abb76634	Ternamyl-
418	39	45.3	1178	8	ABO84648	Ab084648	Human can	491	38	44.2	483	5	ABB76611	Abb76611	Ternamyl-
419	39	45.3	1196	2	AA032017	Aay32017	Mouse cat	492	38	44.2	483	5	ABB76632	Abb76632	Ternamyl-
420	39	45.3	1196	3	ADI38331	Adi38331	Mouse cat	493	38	44.2	483	5	ABB76622	Abb76622	Ternamyl-
421	39	45.3	1196	10	AE68557	Aee68557	Mouse pot	494	38	44.2	483	5	ABB76624	Abb76624	Ternamyl-
422	39	45.3	1203	2	AA043261	Aay43261	Human pat	495	38	44.2	483	5	ABB76616	Abb76616	Ternamyl-
423	39	45.3	1203	2	AA028444	Aay28444	Human ptc	496	38	44.2	483	5	ABB76612	Abb76612	Ternamyl-
424	39	45.3	1203	3	AA092703	Aay92703	Human pat	497	38	44.2	483	5	ABB76613	Abb76613	Ternamyl-
425	39	45.3	1203	3	AA019829	Aae19829	Human pat	498	38	44.2	483	5	ABB76615	Abb76615	Ternamyl-
426	39	45.3	1203	6	ABG74104	Abg74104	Human pat	499	38	44.2	483	5	ABB76631	Abb76631	Ternamyl-
427	39	45.3	1207	8	ABO84646	Ab084646	Human can	500	38	44.2	483	5	ABB76611	Abb76611	Ternamyl-
428	39	45.3	1236	8	ABO84647	Ab084647	Human can	501	38	44.2	483	5	ABB76623	Abb76623	Ternamyl-
429	39	45.3	1323	7	ADD25207	Ad25207	Fertility	502	38	44.2	483	5	ABB76633	Abb76633	Ternamyl-
430	39	45.3	1323	8	ADN61222	Adn61222	Radish nu	503	38	44.2	483	5	ABB76618	Abb76618	Ternamyl-
431	39	45.3	1331	4	AA039048	Aam39048	Human pol	504	38	44.2	483	5	ABB76625	Abb76625	Ternamyl-
432	39	45.3	1516	8	ADQ97109	Adq97109	Mouse can	505	38	44.2	483	5	ABB76629	Abb76629	Ternamyl-
433	39	45.3	1927	7	ADDA48256	Ad48256	Human pro	506	38	44.2	483	5	ABB76617	Abb76617	Ternamyl-
434	39	45.3	1933	8	ABM82906	Abm82906	Human dia	507	38	44.2	483	5	ABB76630	Abb76630	Ternamyl-
435	38.5	44.8	366	8	ADX70942	Adx70942	Plant ful	508	38	44.2	488	4	AA031471	Aab31471	Amino aci
436	38.5	44.8	480	8	ADL99494	Adl99494	Human leu	509	38	44.2	488	4	AA031472	Aab31472	Amino aci
437	38.5	44.8	532	8	ADL99491	Adl99491	Human leu	510	38	44.2	488	4	AA031472	Aab31472	Amino aci
438	38.5	44.8	563	5	ABBA48302	Abb48302	Listeria	511	38	44.2	488	4	AA031470	Aab31470	Amino aci
439	38.5	44.8	601	10	AEF72836	Aef72836	Human leu	512	38	44.2	502	8	ADY11010	Ady11010	Plant ful
440	38.5	44.8	610	8	ADT66633	Adt66633	Rat leuko	513	38	44.2	519	6	ABP98897	Abp98897	Human mol
441	38.5	44.8	611	3	AA080840	Aab080840	Amino aci	514	38	44.2	607	2	AAW09422	Aaw09422	Banana po
442	38.5	44.8	611	7	AD025724	Ad025724	Human pro	515	38	44.2	645	8	ADJ93708	Adj93708	ACH22 pro
443	38.5	44.8	611	8	ABM81460	Abm81460	Tumour-as	516	38	44.2	647	4	ABB60060	Abb60060	Drosophil
444	38.5	44.8	611	9	ADW07254	Adw07254	Human hep	517	38	44.2	649	4	ABB66538	Abb66538	Drosophil
445	38.5	44.8	611	9	ADZ04287	Adz04287	Human leu	518	38	44.2	658	4	ABG27820	Abg27820	Novel hum
446	38.5	44.8	625	3	AA058111	Aab58111	Lung can	519	38	44.2	773	9	AEC08348	Aec08348	A. evansi
447	38	44.2	14	AA012159	Aae12159	Human hg-	520	38	44.2	798	4	AAU03530	Aau03530	Human pro	
448	38	44.2	60	4	ADA81300	Ada81300	Human FAB	521	38	44.2	798	4	AA035330	Aab35330	Human kin
449	38	44.2	128	6	ADAB3800	Adab3800	Human sec	522	38	44.2	804	8	ADX66686	Adx66686	Plant ful
450	38	44.2	140	3	AA032657	Aag032657	Novel hum	523	38	44.2	891	4	ABB64847	Abb64847	Drosophil
451	38	44.2	152	4	AAU32657	Aau32657	Novel hum	524	38	44.2	904	6	ADB17511	Adb17511	Rice post
452	38	44.2	182	6	ABU44226	Abu44226	Protein e	525	38	44.2	904	9	AEC75735	Aec75735	Rice japo
453	38	44.2	185	4	ABG69211	Abg69211	Drosophil	526	38	44.2	915	7	ABO61732	Ab061732	Klebsiell
454	38	44.2	186	4	ABU53118	Abu53118	Intracell	527	38	44.2	916	6	ADB17498	Adb17498	Wheat pos
455	38	44.2	192	8	ADX72427	Adx72427	Plant ful	528	38	44.2	916	9	AEC75722	Aec75722	Wheat Arg
456	38	44.2	195	9	AEC40516	Aec40516	Deltan50	529	38	44.2	932	6	ABU16693	Abu16693	Protein e
457	38	44.2	207	9	AEC40515	Aec40515	Deltan40	530	38	44.2	964	2	AAW32619	Aaw32619	Cyclic-is
458	38	44.2	210	2	AAW23087	Aaw23087	Polyangiu	531	38	44.2	1005	4	ABG11447	Abg11447	Novel hum
459	38	44.2	222	9	AEC40514	Aec40514	Deltan20	532	38	44.2	1139	8	ADR73937	Adr73937	Human man
460	38	44.2	248	8	ADX68808	Adx68808	Plant ful	533	38	44.2	1139	9	ABE71164	Aeb71164	Alpha-man
461	38	44.2	248	9	AEC40497	Aec40497	BFDV isol	534	38	44.2	1180	4	AAU32658	Aau32658	Novel hum

535	38	44.2	1939	4	ABG20505	Abg20505 Novel hum	608	37	43.0	269	9	AEA58368	Aea58368 Streptoco
536	38	44.2	2017	4	AAU32654	Aau32654 Novel hum	609	37	43.0	273	7	ADC18688	Adc18688 Mouse gp
537	38	44.2	2202	4	AAU32655	Aau32655 Novel hum	610	37	43.0	275	7	ADC18689	Adc18689 Mouse gp
538	38	44.2	2858	4	ABB71150	Abb71150 Drosophil	611	37	43.0	276	7	ADC18687	Adc18687 Mouse gp
539	38	44.2	3050	4	ABW58064	Abw58064 Drosophil	612	37	43.0	280	7	ABM86043	Abm86043 Rice abio
540	37.5	43.6	41	2	AAW53674	Aaw53674 FIV PPR c	613	37	43.0	283	8	ADK47844	Adk47844 Streptoco
541	37.5	43.6	288	9	ADX40464	Adx40464 HIV Vif p	614	37	43.0	297	4	ABG15100	Abg15100 Novel hum
542	37.5	43.6	347	8	ABM84252	Abm84252 Human dia	615	37	43.0	310	10	AEF54140	Aef54140 His tagge
543	37.5	43.6	420	8	ABM84251	Abm84251 Human dia	616	37	43.0	315	7	ADD12568	Add12568 Human ENZ
544	37.5	43.6	446	8	ABM84250	Abm84250 Human dia	617	37	43.0	316	7	ABO81579	AbO81579 Pseudomon
545	37.5	43.6	449	7	ADC78253	Adc78253 Human sec	618	37	43.0	319	9	ABE42107	Aeb42107 L. pneumo
546	37.5	43.6	449	9	AED11499	Aed11499 Human gen	619	37	43.0	323	9	ABE38947	Aeb38947 L. pneumo
547	37.5	43.6	546	5	ABB57161	Abb57161 Mouse isc	620	37	43.0	332	8	ADY12827	Ady12827 plant ful
548	37.5	43.6	550	8	ADM33966	Adm33966 Human DGC	621	37	43.0	374	8	ADS27118	Ads27118 Bacterial
549	37.5	43.6	550	9	ADX06171	Adx06171 Cycilin-de	622	37	43.0	374	8	ADS26366	Ads26366 Bacterial
550	37.5	43.6	576	8	ABM84248	Abm84248 Human dia	623	37	43.0	374	8	ADS26734	Ads26734 Bacterial
551	37.5	43.6	616	8	ABM84249	Abm84249 Human dia	624	37	43.0	375	8	ADN21639	Adn21639 Bacterial
552	37	43.0	8	3	AAV67360	Aay67360 Melanoma	625	37	43.0	375	8	ADN24395	Adn24395 Bacterial
553	37	43.0	9	3	AAV67360	Aay67360 Human tyr	626	37	43.0	379	8	ADS43401	Adn43401 Bacterial
554	37	43.0	9	3	AAV67360	Aay67360 Human tyr	627	37	43.0	388	8	ADN21799	Adn21799 Bacterial
555	37	43.0	9	4	AAU26665	Aau26665 Human Leu	628	37	43.0	388	8	ADN24559	Adn24559 Bacterial
556	37	43.0	9	4	AAU26698	Aau26698 Human Leu	629	37	43.0	389	6	ABU21323	Abu21323 Protein e
557	37	43.0	9	4	AAU26697	Aau26697 Human Leu	630	37	43.0	395	6	ABU21149	Abu21149 Protein e
558	37	43.0	9	4	AAU26664	Aau26664 Human Leu	631	37	43.0	408	4	ABG96440	Abg96440 Putative
559	37	43.0	9	5	AAE31324	Aae31324 Human tyr	632	37	43.0	420	7	ABO76672	AbO76672 Pseudomon
560	37	43.0	9	5	AAE31384	Aae31384 Human Tyr	633	37	43.0	431	7	ADY99069	Ady99069 Human KPP
561	37	43.0	9	5	AAE31325	Aae31325 Human tyr	634	37	43.0	443	9	ADY39715	Ady39715 MADS box
562	37	43.0	9	7	ADN14217	Adn14217 Human mel	635	37	43.0	445	8	ADN25882	Adn25882 Bacterial
563	37	43.0	9	7	ADW31578	Adw31578 HLA bindi	636	37	43.0	447	9	AEA81265	Aea81265 Human gal
564	37	43.0	10	4	AAU26718	Aau26718 Human Leu	637	37	43.0	448	3	AAV71285	Aay71285 Streptoco
565	37	43.0	10	4	AAU26707	Aau26707 Human Leu	638	37	43.0	448	6	ABU02732	Abu02732 S. pneumo
566	37	43.0	10	4	AAU27051	Aau27051 Human Leu	639	37	43.0	448	8	ADM92272	Adm92272 S. pneumo
567	37	43.0	10	4	AAU27040	Aau27040 Human Leu	640	37	43.0	448	8	ADT50163	Adt50163 S.pneumon
568	37	43.0	23	2	AAV36451	Aay36451 Fragment	641	37	43.0	448	9	ABE91520	AbE91520 Microbial
569	37	43.0	23	6	ADA11986	Ada11986 Human nov	642	37	43.0	458	9	AEA81266	Aea81266 Human gal
570	37	43.0	27	10	AEF54139	Aef54139 Histidine	643	37	43.0	467	2	AAV35800	Aay35800 Amino aci
571	37	43.0	30	5	AAU84938	Aau84938 Human Trp	644	37	43.0	468	8	ABM83258	Abm83258 Human dia
572	37	43.0	30	5	AAU84873	Aau84873 Human Trp	645	37	43.0	488	8	ABM83259	Abm83259 Human dia
573	37	43.0	32	2	AAR97388	Aar97388 CC49 VH-s	646	37	43.0	491	3	ABAB36459	Abab36459 Lemon acy
574	37	43.0	41	4	ABBI17688	Abbi17688 Human ner	647	37	43.0	499	7	ADG70544	Adg70544 Aspergill
575	37	43.0	51	5	ABP63707	Abp63707 Human ORF	648	37	43.0	500	8	ADS42425	AdS42425 Bacterial
576	37	43.0	52	4	AAU62413	Aau62413 Propionib	649	37	43.0	502	6	ABU63509	Abu63509 Ornithine
577	37	43.0	52	6	ABM58932	Abm58932 Propionib	650	37	43.0	517	2	AAV31981	Aay31981 Mouse tyr
578	37	43.0	53	4	AAU06399	Aau06399 Human foe	651	37	43.0	517	2	AAV42635	Aay42635 Murine ty
579	37	43.0	62	2	AAW20936	Aaw20936 H. pylori	652	37	43.0	517	5	AAU76663	Aau76663 Mouse tyr
580	37	43.0	67	4	AAU46194	Aau46194 Propionib	653	37	43.0	517	6	ABR82196	AbR82196 Tyrosinas
581	37	43.0	67	6	ABM42713	Abm42713 Propionib	654	37	43.0	519	2	AAW30826	Aaw30826 The novel
582	37	43.0	80	4	AAU63600	Aau63600 Propionib	655	37	43.0	519	2	AAV31982	Aay31982 Human tyr
583	37	43.0	80	6	ABM60119	Abm60119 Propionib	656	37	43.0	519	4	AAV42636	Aay42636 Human tyr
584	37	43.0	85	2	AAV36449	Aay36449 Fragment	657	37	43.0	519	4	AAAB86041	Aab86041 Human Trp
585	37	43.0	85	6	ADA11984	Ada11984 Human nov	658	37	43.0	519	5	AAU84807	Aau84807 Human Trp
586	37	43.0	86	4	ADS08142	AdS08142 Staphyloc	659	37	43.0	519	6	ADI28007	Adi28007 Human tyr
587	37	43.0	96	4	AAO12637	Aao12637 Human pol	660	37	43.0	519	6	ABR58675	AbR58675 Human can
588	37	43.0	107	5	ABP03356	Abp03356 Human ORF	661	37	43.0	519	8	ADM12479	Adm12479 Human dop
589	37	43.0	110	2	AAV21366	Aay21366 Human HUP	662	37	43.0	519	8	ADO38702	AdO38702 Tyrosinas
590	37	43.0	111	6	ABR64186	AbR64186 Angiogene	663	37	43.0	519	8	ADQ19864	Adq19864 Human sof
591	37	43.0	116	8	ADX92978	Adx92978 Plant ful	664	37	43.0	519	9	ADY85097	AdY85097 Tumor ant
592	37	43.0	127	2	AAW20574	Aaw20574 H. pylori	665	37	43.0	519	10	AEF01187	Aef01187 Tyrosinas
593	37	43.0	127	7	ADE57168	Ade57168 Human Pro	666	37	43.0	537	4	AAAB86040	Aab86040 Human Trp
594	37	43.0	157	7	ABO74821	AbO74821 Pseudomon	667	37	43.0	537	5	ADI28006	Adi28006 Human tyr
595	37	43.0	181	8	ADN18913	Adn18913 Bacterial	668	37	43.0	546	8	ADN17427	Adn17427 Bacterial
596	37	43.0	181	8	ADN18897	Adn18897 Bacterial	669	37	43.0	551	4	ABAB48829	Aab48829 Thermotog
597	37	43.0	182	3	ABAB29563	Abab29563 Trypanoso	670	37	43.0	552	5	ABO7927	AbO7927 T. mariti
598	37	43.0	198	4	ABG30053	Abg30053 Novel hum	671	37	43.0	552	6	ABB82547	AbB82547 Human tyr
599	37	43.0	223	8	ADV82139	Adv82139 Streptoco	672	37	43.0	552	8	ADRS1311	AdR51311 Anti-biof
600	37	43.0	225	5	ABP25844	AbP25844 Streptoco	673	37	43.0	554	8	ADS23691	AdS23691 Bacterial
601	37	43.0	225	8	ADV88748	Adv88748 Streptoco	674	37	43.0	555	2	AAW34563	Aaw34563 Thermotog
602	37	43.0	225	8	ADV80001	Adv80001 Streptoco	675	37	43.0	555	2	AAW49867	Aaw49867 Thermotog
603	37	43.0	234	2	AAE60794	Aar60794 Kawasaki	676	37	43.0	555	4	ABD73254	AbB73254 Protein r
604	37	43.0	248	4	AAAB24180	Aab24180 Pseudomon	677	37	43.0	555	7	ADC26915	AdC26915 Thermotog
605	37	43.0	248	5	AAU84805	Aau84805 Human TRP	678	37	43.0	555	7	ADE93811	AdE93811 T. mariti
606	37	43.0	249	2	AAAR14258	Aar14258 gp75 pept	679	37	43.0	555	9	ADW95037	AdW95037 Torenia h
607	37	43.0	269	8	ADR94498	Adr94498 Novel S.	680	37	43.0	555	9	ADX05239	AdX05239 Torenia h

681	37	43.0	576	3	AAV75498	Aay75498 Neisseria	754	36	41.9	9	9	ADU70736	AdU70736 Human hep
682	37	43.0	576	6	ABP80808	N. gonorr	755	36	41.9	11	10	AEES9474	AeeS9474 Natriuret
683	37	43.0	618	2	AAW34991	Thermotog	756	36	41.9	13	10	AEES9467	AeeS9467 Natriuret
684	37	43.0	684	4	ABG24375	Novel hum	757	36	41.9	15	9	ADU71289	AdU71289 Human hep
685	37	43.0	751	4	AAE09958	Methylomo	758	36	41.9	17	3	AAU06401	AaU06401 Randomise
686	37	43.0	751	5	ABG61579	High grow	759	36	41.9	48	8	ADJ12358	AdJ12358 Peptide f
687	37	43.0	751	6	ABP77444	N. gonorr	760	36	41.9	61	4	AAU48383	AaU48383 Propionib
688	37	43.0	773	3	AAU08200	Amino aci	761	36	41.9	61	5	ABP01082	AbP01082 Human ORF
689	37	43.0	773	3	ADW23592	Human dia	762	36	41.9	61	6	ABM44902	AbM44902 Propionib
690	37	43.0	773	10	AEF14405	Human cho	763	36	41.9	63	6	ABM65503	AbM65503 Propionib
691	37	43.0	796	6	AAU17006	Protein e	764	36	41.9	65	4	AAU81629	AaU81629 Human hae
692	37	43.0	797	5	ABE19149	Human kin	765	36	41.9	65	4	AAU81629	AaU81629 Human hae
693	37	43.0	804	3	AAU08199	Amino aci	766	36	41.9	97	4	AAU54279	AaU54279 Propionib
694	37	43.0	804	8	ADJ96665	Human lip	767	36	41.9	97	6	ABM50798	AbM50798 Propionib
695	37	43.0	804	10	AEF14404	Human cho	768	36	41.9	100	4	AAU67296	AaU67296 Propionib
696	37	43.0	869	5	ABP65968	Bifidobac	769	36	41.9	100	6	ABM63815	AbM63815 Propionib
697	37	43.0	952	4	ABB69760	Drosophil	770	36	41.9	100	7	ABO63532	AbO63532 Klebsiell
698	37	43.0	1023	4	ABG24378	Novel hum	771	36	41.9	100	7	AAU51623	AaU51623 Propionib
699	37	43.0	1032	6	ABU15004	Protein e	772	36	41.9	113	4	AAU51623	AaU51623 Propionib
700	37	43.0	1043	4	ABBS8772	Drosophil	773	36	41.9	113	6	ABM48142	AbM48142 Propionib
701	37	43.0	1156	8	ADS42937	Bacterial	774	36	41.9	113	6	ABJ25288	AbJ25288 Mouse BAC
702	37	43.0	1201	8	ADS30065	Bacterial	775	36	41.9	115	8	ADP30206	AdP30206 Human sec
703	37	43.0	1368	5	ABP73707	Candida a	776	36	41.9	116	4	AAU63310	AaU63310 Propionib
704	37	43.0	1368	5	ADJ28002	Human TRP	777	36	41.9	116	6	ABM59829	AbM59829 Propionib
705	37	43.0	1384	5	ADJ28004	Human TRP	778	36	41.9	121	3	AAU20409	AaU20409 Arabidops
706	37	43.0	1392	5	ADJ28000	Human TRP	779	36	41.9	122	5	ABP11305	AbP11305 Human ORF
707	37	43.0	1605	4	ABBT70375	Drosophil	780	36	41.9	131	8	ADK71682	AdK71682 pHSN-M6SL
708	37	43.0	1674	9	AEBS2335	Human pro	781	36	41.9	135	7	ABO78076	AbO78076 Pseudomon
709	37	43.0	1674	9	AEBS2340	Human pro	782	36	41.9	158	5	ABB97984	AbB97984 Human cau
710	37	43.0	1793	4	ABG25611	Novel hum	783	36	41.9	158	5	ADK71841	AdK71841 Human kin
711	37	43.0	2071	4	ABU04312	Human exp	784	36	41.9	168	8	ABU34183	AbU34183 Protein e
712	37	43.0	2338	6	ABU04315	Human exp	785	36	41.9	178	6	ABU52835	AbU52835 Human sig
713	37	43.0	2491	5	AAE24315	Human cat	786	36	41.9	184	4	ABU52835	AbU52835 Human sig
714	37	43.0	2491	5	ABU04319	Human exp	787	36	41.9	184	4	ABU53120	AbU53120 Intracell
715	37	43.0	2491	6	ABU04316	Human exp	788	36	41.9	184	4	ABU53120	AbU53120 Intracell
716	37	43.0	2491	6	ABU04313	Human exp	789	36	41.9	187	7	ABO73380	AbO73380 Pseudomon
717	37	43.0	2491	6	ABU04317	Human exp	790	36	41.9	187	7	ABO73380	AbO73380 Pseudomon
718	37	43.0	2491	6	ABU04316	Human exp	791	36	41.9	198	3	AAU20408	AaU20408 Arabidops
719	37	43.0	2491	6	ABU04313	Human exp	792	36	41.9	198	3	AAU20408	AaU20408 Arabidops
720	37	43.0	2491	6	ABU04313	Human exp	793	36	41.9	198	8	ADT55527	AdT55527 Plant pol
721	37	43.0	2491	6	ABU04320	Human exp	794	36	41.9	198	8	ADT55527	AdT55527 Plant pol
722	37	43.0	2491	7	ADJ68259	Human hea	795	36	41.9	198	8	ADT55527	AdT55527 Plant pol
723	37	43.0	2491	8	ADJ68259	Human hea	796	36	41.9	203	4	AAU20408	AaU20408 Arabidops
724	37	43.0	2491	8	ADJ68259	Human hea	797	36	41.9	203	4	AAU20408	AaU20408 Arabidops
725	37	43.0	2491	8	ADJ68259	Human hea	798	36	41.9	203	9	ADU09536	AdU09536 Human pro
726	37	43.0	2491	8	ADJ68259	Human hea	799	36	41.9	211	8	ADU09536	AdU09536 Human pro
727	37	43.0	2491	9	ADJ68259	Human hea	800	36	41.9	211	8	ADU09536	AdU09536 Human pro
728	37	43.0	2491	9	ADJ68259	Human hea	801	36	41.9	219	4	AAU62002	AaU62002 Propionib
729	37	43.0	2491	9	ADJ68259	Human hea	802	36	41.9	219	4	AAU62002	AaU62002 Propionib
730	37	43.0	2491	9	ADJ68259	Human hea	803	36	41.9	219	4	AAU62002	AaU62002 Propionib
731	37	43.0	2681	4	AAW61348	Mouse tel	804	36	41.9	219	4	AAU62002	AaU62002 Propionib
732	37	43.0	2681	4	AAW629184	Novel hum	805	36	41.9	225	8	ADY06493	AdY06493 Plant ful
733	37	43.0	2824	8	ADN22513	Bacterial	806	36	41.9	225	8	ADY06493	AdY06493 Plant ful
734	37	43.0	5546	5	AAU85008	Human mel	807	36	41.9	226	4	AAU20408	AaU20408 Arabidops
735	36.5	42.4	7026	10	AEF57851	Polyketid	808	36	41.9	226	4	AAU20408	AaU20408 Arabidops
736	36.5	42.4	101	4	AAO03669	Human pol	809	36	41.9	242	6	ABU23340	AbU23340 Protein e
737	36.5	42.4	157	6	ABU09576	VEE virus	810	36	41.9	248	3	AAU23340	AaU23340 Protein e
738	36.5	42.4	165	6	ABU09577	Mutant VE	811	36	41.9	248	3	AAU23340	AaU23340 Protein e
739	36.5	42.4	165	6	ABU09578	Mutant VE	812	36	41.9	251	8	ADX78625	AdX78625 Plant ful
740	36.5	42.4	169	6	ABU09579	Mutant VE	813	36	41.9	251	8	ADX78625	AdX78625 Plant ful
741	36.5	42.4	191	10	AEF63481	Modified	814	36	41.9	251	8	ADX78625	AdX78625 Plant ful
742	36.5	42.4	192	8	ADP20069	Human imm	815	36	41.9	253	3	AAU33386	AaU33386 Zea may
743	36.5	42.4	192	9	ADX40511	HIV Vif p	816	36	41.9	253	3	AAU33386	AaU33386 Zea may
744	36.5	42.4	192	9	ADX40520	HIV Vif p	817	36	41.9	254	8	ADS23763	AdS23763 Bacterial
745	36.5	42.4	192	9	ADX40512	HIV Vif p	818	36	41.9	254	8	ADS23763	AdS23763 Bacterial
746	36.5	42.4	192	9	ADX40510	HIV Vif p	819	36	41.9	259	5	ABU51611	AbU51611 Helicobac
747	36.5	42.4	268	5	ADP25596	Streptoco	820	36	41.9	259	5	ABU51611	AbU51611 Helicobac
748	36.5	42.4	268	8	ADR83944	S. pyogen	821	36	41.9	283	6	ABU19995	AbU19995 Protein e
749	36.5	42.4	288	9	ADX40444	HIV Vif p	822	36	41.9	283	6	ABU19995	AbU19995 Protein e
750	36.5	42.4	342	5	AAE18305	Venezuela	823	36	41.9	283	6	ABU19995	AbU19995 Protein e
751	36.5	42.4	381	7	ABM88307	Rice abio	824	36	41.9	296	7	ABO63702	AbO63702 Klebsiell
752	36.5	42.4	696	7	ABO79890	Pseudomon	825	36	41.9	296	7	ABO63702	AbO63702 Klebsiell
753	36.5	42.4	1439	10	ABF63466	pol-linke	826	36	41.9	301	5	ABG91590	AbG91590 Purine/py
						827	36	41.9	301	8	ADS25903	AdS25903 Bacterial	
						828	36	41.9	301	8	ADS25903	AdS25903 Bacterial	
						829	36	41.9	302	9	AEC34213	AeC34213 Baculovir	
						830	36	41.9	302	9	AEC34213	AeC34213 Baculovir	
						831	36	41.9	317	3	ABG35263	AbG35263 Zea may	
						832	36	41.9	317	3	ABG35263	AbG35263 Zea may	
						833	36	41.9	320	5	ABG91593	AbG91593 Purine/py	
						834	36	41.9	320	5	ABG91593	AbG91593 Purine/py	
						835	36	41.9	321	5	AAU75802	AaU75802 Rice morp	
						836	36	41.9	321	5	AAU75802	AaU75802 Rice morp	
						837	36	41.9	321	10	AEF29823	Aef29823 Lead Cere	
						838	36	41.9	327	3	AAU75802	AaU75802 Rice morp	
						839	36	41.9	327	3	AAU75802	AaU75802 Rice morp	
						840	36	41.9	336	6	ABU30722	AbU30722 Protein e	
						841	36	41.9	336	6	ABU30722	AbU30722 Protein e	
						842	36	41.9	344	3	AAU35261	AaU35261 Zea may	
						843	36	41.9	344	3	AAU35261	AaU35261 Zea may	
						844	36	41.9	356	8	ADS43881	AdS43881 Bacterial	
						845	36	41.9	356	8	ADS43881	AdS43881 Bacterial	
						846	36	41.9	358	8	ADJ12357	AdJ12357 Peptide f	
						847	36	41.9	358	8	ADJ12357	AdJ12357 Peptide f	
						848	36	41.9	370	7	ADU7199	AdU7199 Novel sig	
						849	36	41.9	370	7	ADU7199	AdU7199 Novel sig	
						850	36	41.9	370	7	ADU7199	AdU7199 Novel sig	
						851	36	41.9	372	7	ADU7199	AdU7199 Novel sig	
						852	36	41.9	372	7	ADU7199	AdU7199 Novel sig	
						853	36	41.9	382	8	ADT59274	AdT59274 Plant pol	
						854	36	41.9	382	8	ADT59274	AdT59274 Plant pol	

827	36	41.9	383	9	ABE37429	Aeb37429	L. pneumo	900	36	41.9	485	3	AAB29363	Aab29363	Bacillus
828	36	41.9	385	6	ABU17339	Abu17339	Protein e	901	36	41.9	485	3	AAB29387	Aab29387	Bacillus
829	36	41.9	391	6	ADA33574	Ada33574	Acinetoba	902	36	41.9	485	3	AAB29389	Aab29389	Bacillus
830	36	41.9	396	8	ADY13837	Ady13837	Plant ful	903	36	41.9	485	3	AAB29385	Aab29385	Bacillus
831	36	41.9	423	8	ADN22904	Adn22904	Bacterial	904	36	41.9	485	3	AAB29365	Aab29365	Bacillus
832	36	41.9	426	7	ABM65116	Abm65116	Propionib	905	36	41.9	485	3	AAB29386	Aab29386	Bacillus
833	36	41.9	429	6	ABM88474	Abm88474	Rice abio	906	36	41.9	485	3	AAB29388	Aab29388	Bacillus
834	36	41.9	430	6	ADA54609	Ada54609	Human pro	907	36	41.9	485	3	AAB29394	Aab29394	Bacillus
835	36	41.9	441	8	ABY00045	AbY00045	Porcine e	908	36	41.9	485	3	AAB29366	Aab29366	Bacillus
836	36	41.9	443	4	ABG14210	Abg14210	Novel hum	909	36	41.9	488	9	AED19768	Aed19768	Baculovir
837	36	41.9	447	6	ADA55102	Ada55102	Human pro	910	36	41.9	491	9	AED19774	Aed19774	Baculovir
838	36	41.9	449	6	ABU18372	Abu18372	Protein e	911	36	41.9	502	3	AAG06850	Aag06850	Arabidops
839	36	41.9	450	3	AAG41505	Aag41505	Arabidops	912	36	41.9	511	2	AAW78475	Aaw78475	Autograph
840	36	41.9	450	5	ABB93654	Abb93654	Herbicida	913	36	41.9	512	8	ADG16321	Adg16321	Antibody
841	36	41.9	452	3	AAG06851	Aag06851	Arabidops	914	36	41.9	512	8	AED19766	Aed19766	Baculovir
842	36	41.9	453	8	ADN20522	Adn20522	Bacterial	915	36	41.9	517	5	AAE25051	Aae25051	GP64-6His
843	36	41.9	455	3	AAG41504	Aag41504	Arabidops	916	36	41.9	518	8	ADG16323	Adg16323	Antibody
844	36	41.9	458	3	AAG41503	Aag41503	Arabidops	917	36	41.9	518	8	ADT04559	Adt04559	Peptide t
845	36	41.9	463	3	AAG06588	Aag06588	Arabidops	918	36	41.9	528	8	ABM83086	Abm83086	Human dia
846	36	41.9	463	3	AAG48388	Aag48388	Arabidops	919	36	41.9	530	2	AAW78476	Aaw78476	Baculovir
847	36	41.9	463	5	ABU93655	Abu93655	Herbicida	920	36	41.9	532	2	AAI39296	Aai39296	Alkaline
848	36	41.9	463	5	ABU36773	Abu36773	Protein e	921	36	41.9	532	8	ABM83085	Abm83085	Human dia
849	36	41.9	470	3	AAG06587	Aag06587	Arabidops	922	36	41.9	538	9	ADV14310	Adv14310	Corn Na+/
850	36	41.9	470	3	AAG48387	Aag48387	Arabidops	923	36	41.9	538	9	ADW95062	Adw95062	Zea maye
851	36	41.9	481	2	AAW55029	Aaw55029	G-protein	924	36	41.9	538	9	ADX05264	Adx05264	Zea maye
852	36	41.9	481	2	AAI25969	Aai25969	Human bra	925	36	41.9	540	4	ABB64912	Abb64912	Drosophi
853	36	41.9	481	2	AAI25969	Aai25969	Human bra	926	36	41.9	542	2	AAW55030	Aaw55030	G-protein
854	36	41.9	481	3	AAI25969	Aai25969	Human bra	927	36	41.9	545	8	ABM83087	Abm83087	Human dia
855	36	41.9	481	3	AAI25969	Aai25969	Human bra	928	36	41.9	545	8	ADU07866	Adu07866	Amino aci
856	36	41.9	481	3	AAI25969	Aai25969	Human bra	929	36	41.9	576	8	ABM83084	Abm83084	Human dia
857	36	41.9	481	5	ABG95173	Abg95173	Human GPC	930	36	41.9	587	4	ABM67655	Abm67655	Amino aci
858	36	41.9	481	5	ABG95160	Abg95160	Human GPC	931	36	41.9	587	8	ABM83083	Abm83083	Human dia
859	36	41.9	481	6	ABP82006	Abp82006	Human end	932	36	41.9	587	8	ADU07900	Adu07900	Amino aci
860	36	41.9	481	6	ABP82006	Abp82006	Human end	933	36	41.9	587	8	ADU07904	Adu07904	Amino aci
861	36	41.9	481	7	ADB57657	Adb57657	Human end	934	36	41.9	587	8	ADW76538	Adw76538	Alkaliph
862	36	41.9	481	7	ADC22601	Adc22601	Human G p	935	36	41.9	590	8	ADU07894	Adu07894	Amino aci
863	36	41.9	481	7	ADC22727	Adc22727	Human G p	936	36	41.9	590	8	ADU07912	Adu07912	Amino aci
864	36	41.9	481	7	ADH14200	Adh14200	Mutated h	937	36	41.9	591	8	ABM83082	Abm83082	Human dia
865	36	41.9	481	7	ADH14074	Adh14074	Human ETB	938	36	41.9	601	8	ABM83081	Abm83081	Human dia
866	36	41.9	481	8	ADO29058	Ado29058	Human nov	939	36	41.9	616	8	ABM83089	Abm83089	Human dia
867	36	41.9	481	8	ABM80229	Abm80229	Tumour-as	940	36	41.9	671	6	ABR52866	AbR52866	Protein s
868	36	41.9	481	9	AED34533	Aed34533	Human ETB	941	36	41.9	671	7	ADK62338	Adk62338	Disease t
869	36	41.9	482	3	ADO29059	Ado29059	Mouse nov	942	36	41.9	671	8	ADN19334	Adn19334	Bacterial
870	36	41.9	483	3	AAB29370	Aab29370	Bacillus	943	36	41.9	672	8	ABM83080	Abm83080	Human dia
871	36	41.9	483	3	AAB29372	Aab29372	Bacillus	944	36	41.9	683	8	ADJ75550	Adj75550	Marker ge
872	36	41.9	483	3	AAB29380	Aab29380	Bacillus	945	36	41.9	685	8	ADR39739	Adr39739	Human kin
873	36	41.9	483	3	AAB29371	Aab29371	Bacillus	946	36	41.9	698	4	ABG08662	Abg08662	Novel hum
874	36	41.9	483	3	AAB29376	Aab29376	Bacillus	947	36	41.9	711	8	ABM83079	Abm83079	Human dia
875	36	41.9	483	3	AAB29325	Aab29325	Bacillus	948	36	41.9	715	6	ABU03523	Abu03523	Angiogene
876	36	41.9	483	3	AAB29384	Aab29384	Bacillus	949	36	41.9	715	8	ADN03820	Adn03820	Antipeori
877	36	41.9	483	3	AAB29328	Aab29328	Bacillus	950	36	41.9	715	8	ADO19246	Ado19246	Human PRO
878	36	41.9	483	3	AAB29377	Aab29377	Bacillus	951	36	41.9	715	8	ABM83078	Abm83078	Human dia
879	36	41.9	483	3	AAB29378	Aab29378	Bacillus	952	36	41.9	715	8	ABM83077	Abm83077	Human dia
880	36	41.9	483	3	AAB29382	Aab29382	Bacillus	953	36	41.9	716	8	ABM83076	Abm83076	Human dia
881	36	41.9	483	3	AAB29374	Aab29374	Bacillus	954	36	41.9	719	7	ADF70494	Adf70494	Orphan re
882	36	41.9	483	3	AAB29379	Aab29379	Bacillus	955	36	41.9	735	7	ADE76568	Ade76568	Adeno-ass
883	36	41.9	483	3	AAB29369	Aab29369	Bacillus	956	36	41.9	735	7	ADE76569	Ade76569	Adeno-ass
884	36	41.9	483	3	AAB29373	Aab29373	Bacillus	957	36	41.9	735	7	ADE76567	Ade76567	Adeno-ass
885	36	41.9	483	3	AAB29381	Aab29381	Bacillus	958	36	41.9	735	9	ADZ27084	Adz27084	Adeno-ass
886	36	41.9	483	3	AAB29375	Aab29375	Bacillus	959	36	41.9	740	8	ABM83075	Abm83075	Human dia
887	36	41.9	483	3	AAB29329	Aab29329	Bacillus	960	36	41.9	740	8	ABM83074	Abm83074	Human dia
888	36	41.9	483	3	AAB29330	Aab29330	Bacillus	961	36	41.9	759	8	ABM83088	Abm83088	Human dia
889	36	41.9	483	3	AAB29383	Aab29383	Bacillus	962	36	41.9	784	8	ADS30331	Ads30331	Bacterial
890	36	41.9	484	6	ABU34698	Abu34698	Protein e	963	36	41.9	785	9	ABM90808	Abm90808	M. xanthu
891	36	41.9	485	3	AAB29364	Aab29364	Bacillus	964	36	41.9	840	9	ADM00948	Adm00948	Amino aci
892	36	41.9	485	3	AAB29392	Aab29392	Bacillus	965	36	41.9	845	8	ADY23636	Ady23636	Plant ful
893	36	41.9	485	3	AAB29327	Aab29327	Bacillus	966	36	41.9	907	8	ADP29998	Adp29998	Human sec
894	36	41.9	485	3	AAB29390	Aab29390	Bacillus	967	36	41.9	912	4	AAU52873	Aau52873	Propionib
895	36	41.9	485	3	AAB29393	Aab29393	Bacillus	968	36	41.9	912	6	ABM49392	Abm49392	Propionib
896	36	41.9	485	3	AAB29368	Aab29368	Bacillus	969	36	41.9	914	2	AAI15785	Aai15785	B.thuring
897	36	41.9	485	3	AAB29391	Aab29391	Bacillus	970	36	41.9	924	4	ABE67489	AbE67489	Drosophal
898	36	41.9	485	3	AAB29367	Aab29367	Bacillus	971	36	41.9	956	2	AAI15784	Aai15784	B.thuring
899	36	41.9	485	3	AAB29311	Aab29311	Bacillus	972	36	41.9	1100	2	AAI15783	Aai15783	B.thuring

973 36 41.9 1327 8 ADL98347
 974 36 41.9 1413 4 ABB60857
 975 36 41.9 1424 4 ABB60854
 976 36 41.9 1510 7 ADG30698
 977 36 41.9 1536 7 ADK64802
 978 36 41.9 1665 7 ADF04191
 979 36 41.9 1856 3 AAY21805
 980 36 41.9 1856 3 AAY83274
 981 36 41.9 2022 6 ABR63232
 982 36 41.9 3431 8 ADI39260
 983 36 41.9 3723 8 ADU47672
 984 36 41.9 4970 9 AEB00354
 985 36 41.9 36946 9 ADV97835
 986 35.5 41.3 50 2 AAW94136
 987 35.5 41.3 96 2 AAW94134
 988 35.5 41.3 178 6 ABU17135
 989 35.5 41.3 186 6 ADA33747
 990 35.5 41.3 191 9 ADX40531
 991 35.5 41.3 192 1 AAP61504
 992 35.5 41.3 192 2 AAW90177
 993 35.5 41.3 192 3 AAW89324
 994 35.5 41.3 192 3 AAB69298
 995 35.5 41.3 192 4 AAB85994
 996 35.5 41.3 192 6 ABU63326
 997 35.5 41.3 192 6 ABU63321
 998 35.5 41.3 192 8 ADO26433
 999 35.5 41.3 192 8 ADN36415
 1000 35.5 41.3 192 9 ADX40506

ALIGNMENTS

RESULT 1
 ADR88216
 ID ADR88216 standard; peptide; 14 AA.
 AC ADR88216;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human heparanase epitope pep10.
 XX
 KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
 KW human; heparanase; enzyme; epitope.
 XX
 OS Homo sapiens.
 XX
 XX US2004170631-A1.
 XX
 PD 02-SEP-2004.
 XX
 PF 28-NOV-2003; 2003US-00722502.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 PR 22-AUG-2003; 2003US-00645659.
 XX
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (VGEL/) VAN GELDER J. M.
 PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 PI Friedmann Y;
 XX
 XX WPI; 2004-625084/60.
 XX
 PT Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.
 XX
 PS Claim 7; SEQ ID NO 10; 58pp; English.
 CC The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is human heparanase epitope.
 XX
 SQ Sequence 14 AA;
 Query Match 100.0%; Score 86; DB 8; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 |||||
 DB 1 TWHYYLNGRTATR 14
 RESULT 2
 ADR78183
 ID ADR78183 standard; peptide; 14 AA.
 XX
 XX ADR78183;
 AC
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE Functional peptide epitope of human heparanase, pep10.
 XX
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion; disorder;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytotatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
 OS Homo sapiens.

XX US2004213789-A1.
 XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.
 XX 02-SEP-1997; 97US-00922170.
 XX 01-MAY-1998; 98US-00071739.
 XX 04-NOV-1998; 98US-00186200.
 XX 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.
 XX (PERE/) PERETZ T.
 XX (MIRO/) MIRON D.
 XX (SHLO/) SHLOMI Y.
 XX (PECK/) PECKER I.
 XX (AYAL/) AYAL-HERSHKOVITZ M.
 XX (FEIN/) FEINSTEIN E.
 XX (GELD/) GELDER J M V.
 XX (VLOD/) VLADAVSKI I.
 XX (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;
 PI Friedmann Y;
 XX WPI; 2004-774790/76.
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 67; SEQ ID NO 10; 69pp; English.
 XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents a functional peptide
 CC epitope of human heparanase.

XX Sequence 14 AA;

Query Match 100.0%; Score 86; DB 8; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07; Gaps 0;
 Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 TWHYYLNGRTATR 14
 DB 1 TWHYYLNGRTATR 14

RESULT 3
 ID AEA42432

XX AEA42432 standard; peptide; 14 AA.

XX AEA42432;

XX 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:10.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

OS Homo sapiens.

XX AU2004201462-A1.

XX 06-MAY-2004.

XX 08-APR-2004; 2004AU-00201462.

XX 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.

XX Claim 7; SEQ ID NO 10; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase-
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents a human
 CC heparanase epitope peptide, which is used in the exemplification of the
 CC present invention.

XX Sequence 14 AA;

Query Match 100.0%; Score 86; DB 9; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07; Gaps 0;
 Matches 14; Conservative 0; Mismatches 0; Indels 0;

QY 1 TWHYYLNGRTATR 14

PT Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase
 PT activation.

XX Claim 56; SEQ ID NO 21; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular
 CC diseases, neurological diseases or viral infections. This sequence
 CC represents a heparanase inhibitor peptide used in the scope of the
 CC invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 86; DB 9; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTATR 14
 |||||
 Db 2 TWHHYLLNGRTATR 15

RESULT 7

ID ADR88207 standard; protein; 386 AA.

XX AC ADR88207;

XX 18-NOV-2004 (first entry)

DE Human mature heparanase 45 kDa major subunit.

XX Targeted drug delivery ; inflammatory disorder; wound; scar;
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;
 KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.

XX Homo sapiens.

XX US2004170631-A1.

XX 02-SEP-2004.

XX 28-NOV-2003; 2003US-00722502.

XX 02-SEP-1997; 97US-00922170.

XX 01-MAY-1998; 98US-00071739.

XX 04-NOV-1998; 98US-00186200.

XX 19-FEB-2003; 2003US-00368044.

XX 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (VGEL/) VAN GELDER J M.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-HersHKovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 PI Friedmann Y;

XX WPI; 2004-625084/60.

XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.

XX Claim 2; SEQ ID NO 1; 58pp; English.

XX The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature
 CC heparanase.

XX Sequence 386 AA;

Query Match 100.0%; Score 86; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTATR 14
 |||||
 Db 137 TWHHYLLNGRTATR 150

RESULT 8

ADT78174

ID ADT78174 standard; protein; 386 AA.

XX AC ADT78174;

XX 13-JAN-2005 (first entry)

DE 45kDa subunit of mature processed human heparanase dimer.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition ; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cytostatic; antiinflammatory; vulnerary; antiarteriosclerotic;
KW vasototropic; immunosuppressive; nephrotropic; antidiabetic; human.
XX Homo sapiens.

XX US2004213789-A1.

XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (GELD/) GELDER J M V.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX Claim 5; SEQ ID NO 1; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of
CC mature processed human heparanase dimer.

XX Sequence 386 AA;

XX Query Match 100.0%; Score 86; DB 8; Length 386;

Best Local Similarity 100.0%; Pred. No. 6e-06; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 1 TWHHYLNGRTATR 14
DB 137 TWHHYLNGRTATR 150

RESULT 9
ADY27057
ID ADY27057 standard; protein; 386 AA.
XX
AC ADY27057;
XX
DT 05-MAY-2005 (first entry)
XX
DE Heparanase inhibitor protein #1.
XX
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.
XX
OS Homo sapiens.
XX
XX WO2005016227-A2.
XX
XX 24-FEB-2005.
XX
XX 12-AUG-2004; 2004WO-IL000744.
XX
XX 14-AUG-2003; 2003US-0494800P.
XX
XX 12-JAN-2004; 2004US-0535492P.
XX
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX
XX Van-Gelder JM, Miron D;
XX
XX WPI; 2005-182203/19.

PT Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.

XX Claim 55; SEQ ID NO 33; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the
CC invention.

XX Sequence 386 AA;

XX Query Match 100.0%; Score 86; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 6e-06; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 1 TWHYYLNGRTATR 14
 |||||
 Db 137 TWHYYLNGRTATR 150

RESULT 10
 ADZ18995
 ID ADZ18995 standard; protein; 386 AA.
 XX
 AC ADZ18995;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Human heparanase consensus cleavage site #2.
 XX
 KW Enzyme engineering; heparanase; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 PA
 XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
 XX WPI; 2005-273382/28.
 DR
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Disclosure; SEQ ID NO 16; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a human heparanase
 CC consensus cleavage site used in the scope of the invention.
 XX
 SQ Sequence 386 AA;

Query Match 100.0%; Score 86; DB 9; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 |||||
 Db 137 TWHYYLNGRTATR 150

RESULT 11
 AEA42423
 ID AEA42423 standard; protein; 386 AA.
 XX
 AC AEA42423;
 XX
 DT 28-JUL-2005 (first entry)
 XX
 DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.
 XX

KW antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis.
 XX
 OS Homo sapiens.
 XX
 PN AU2004201462-A1.
 XX
 PD 06-MAY-2004.
 XX
 PF 08-APR-2004; 2004AU-00201462.
 XX
 PR 08-APR-2004; 2004AU-00201462.
 XX
 XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 XX
 XX Vlodayevy I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
 PI Feinstein E;
 XX
 DR WPI; 2005-173343/19.
 XX
 XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.
 XX
 PS Claim 2; SEQ ID NO 1; 260pp; English.
 XX
 CC The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents the 45 kDa
 CC subunit of the human mature processed heparanase dimer, which is used in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 386 AA;

Query Match 100.0%; Score 86; DB 9; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 |||||
 Db 137 TWHYYLNGRTATR 150

RESULT 12
 ADY27061
 ID ADY27061 standard; protein; 460 AA.
 XX
 AC ADY27061;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 XX

DE Heparanase inhibitor protein #4.
 XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW heparanase modulator; enzyme purification.
 XX Homo sapiens.
 OS
 XX
 PN WO2005016227-A2.
 XX
 PD 24-FEB-2005.
 XX
 PF 12-AUG-2004; 2004WO-IL000744.
 XX
 PR 14-AUG-2003; 2003US-0494800P.
 PR 12-JAN-2004; 2004US-0535492P.
 XX
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
 XX
 PI Van-Gelder JM, Miron D;
 XX
 DR WPI; 2005-182203/19.
 XX
 PT Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.
 XX
 PS Disclosure; SEQ ID NO 37; 211pp; English.
 XX
 CC The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular
 CC diseases, neurological diseases or viral infections. This sequence
 CC represents a heparanase inhibitor protein used in the scope of the
 CC invention.
 XX
 SQ Sequence 460 AA;
 Query Match 100.0%; Score 86; DB 9; Length 460;
 Best Local Similarity 100.0%; Pred. No. 7.3e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 DB 211 TWHYYLNGRTATR 224
 RESULT 13
 AEB87589
 ID AEB87589 standard; protein; 486 AA.
 XX
 AC AEB87589;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE Human heparanase 65delta20 deletion mutant.
 XX
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;

KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2005071070-A2.
 XX
 PD 04-AUG-2005.
 XX
 PF 20-JAN-2005; 2005WO-IL000068.
 XX
 PR 22-JAN-2004; 2004IL-00160025.
 PR 28-JUL-2004; 2004US-00901943.
 XX
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 XX
 PI Vlodavsky I, Ilan N, Levy-Adam F;
 XX
 DR WPI; 2005-564219/57.
 DR N-PSDB; AEB87588.
 XX
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.
 XX
 PS Claim 107; SEQ ID NO 31; 167pp; English.
 XX
 CC The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX
 SQ Sequence 486 AA;
 Query Match 100.0%; Score 86; DB 9; Length 486;
 Best Local Similarity 100.0%; Pred. No. 7.8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 DB 259 TWHYYLNGRTATR 272
 RESULT 14
 ADZ18996
 ID ADZ18996 standard; protein; 492 AA.
 XX
 AC ADZ18996;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Hep106 construct protein.
 XX
 KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;

KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 XX 17-SEP-2004; 2004WO-EP010517.
 PF
 XX 26-SEP-2003; 2003US-0506479P.
 PR
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX
 XX Lahm A, Nardella C, Pallao M, Steinkuhler C;
 PI
 XX WPI; 2005-273382/28.
 DR
 DR N-PSDB; ADZ18997.
 XX
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 XX Example 2; SEQ ID NO 17; 65pp; English.
 PS
 XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepl06 construct protein
 CC used in the scope of the invention.
 XX
 SQ Sequence 492 AA;
 Query Match 100.0%; Score 86; DB 9; Length 492;
 Best Local Similarity 100.0%; Pred. No. 7.9e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 |||||
 Db 243 TWHYYLNGRTATR 256
 |||||
 RESULT 15
 AEB87562
 ID AEB87562 standard; protein; 493 AA.
 XX
 AC AEB87562;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE Human heparanase 65delta15 deletion mutant.
 XX
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2005071070-A2.
 XX
 PD 04-AUG-2005.
 PR
 XX 20-JAN-2005; 2005WO-IL000068.

XX 22-JAN-2004; 2004IL-00160025.
 PR 28-JUL-2004; 2004US-00901943.
 XX
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 XX
 PI Vlodevsky I, Ilan N, Levy-Adam F;
 XX
 DR WPI; 2005-564219/57.
 DR N-PSDB; AEB87561.
 XX
 XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.
 XX
 PS Claim 105; SEQ ID NO 4; 167pp; English.
 XX
 CC The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX
 SQ Sequence 493 AA;
 Query Match 100.0%; Score 86; DB 9; Length 493;
 Best Local Similarity 100.0%; Pred. No. 7.9e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 |||||
 Db 244 TWHYYLNGRTATR 257
 |||||
 RESULT 16
 ADZ18999
 ID ADZ18999 standard; protein; 495 AA.
 XX
 AC ADZ18999;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Hep109 construct protein.
 XX
 KW Enzyme engineering; heparanase; hepl09; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 XX 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX

PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 XX WPI; 2005-273382/28.
 DR N-PSDB; ADZ18998.
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX Example 2; SEQ ID NO 20; 65pp; English.
 XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoprotease. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepi09 construct protein
 CC used in the scope of the invention.
 XX Sequence 495 AA;
 SQ

Query Match 100.0%; Score 86; DB 9; Length 495;
 Best Local Similarity 100.0%; Pred. NO. 7.9e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 |||||
 DB 246 TWHYYLNGRTATR 259

RESULT 17
 AEB87587
 ID AEB87587 standard; protein; 497 AA.
 AC AEB87587;
 XX
 XX 06-OCT-2005 (first entry)
 DT
 DE Human heparanase 65delta10 deletion mutant.
 XX
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2005071070-A2.
 XX
 PD 04-AUG-2005.
 XX
 XX 20-JAN-2005; 2005WO-IL000068.
 XX
 XX 22-JAN-2004; 2004IL-00160025.
 PR
 PR 28-JUL-2004; 2004US-00901943.
 XX
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 PA
 XX Vlodavsky I, Ilan N, Levy-Adam F;
 PI
 XX WPI; 2005-564219/57.
 DR N-PSDB; AEB87586.
 DR
 XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.

XX Claim 106; SEQ ID NO 29; 167pp; English.
 XX The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX
 SQ Sequence 497 AA;
 Query Match 100.0%; Score 86; DB 9; Length 497;
 Best Local Similarity 100.0%; Pred. NO. 8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 |||||
 DB 248 TWHYYLNGRTATR 261

RESULT 18
 ADZ19000
 ID ADZ19000 standard; protein; 501 AA.
 XX
 AC ADZ19000;
 XX
 XX 16-JUN-2005 (first entry)
 DT
 DE HepGS3 construct protein.
 XX
 KW Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 XX 07-APR-2005.
 XX
 XX 17-SEP-2004; 2004WO-EP010517.
 XX
 XX 26-SEP-2003; 2003US-0506479P.
 PR
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 PA
 XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 PI
 XX WPI; 2005-273382/28.
 DR N-PSDB; ADZ19001.
 DR
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX Example 2; SEQ ID NO 21; 65pp; English.
 PS
 XX

CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepgS3 construct protein
CC used in the scope of the invention.

XX SQ Sequence 501 AA;

Query Match 100.0%; Score 86; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 252 TWHYYLNGRTATR 265
|||||

RESULT 19
ADZ19005
ID ADZ19005 standard; protein; 507 AA.

XX AC ADZ19005;

XX DT 16-JUN-2005 (first entry)

XX DE HepG6 construct protein.

XX KW Enzyme engineering; heparanase; hepgS6; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

XX DR WPI; 2005-273382/28.

XX DR N-PSDB; ADZ19003.

XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.

XX Example 2; SEQ ID NO 26; 65pp; English.

XX PS The invention relates to a synthetic nucleic acid molecule that encodes
PS mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepgS6 construct protein
CC used in the scope of the invention.

XX SQ Sequence 507 AA;

Query Match 100.0%; Score 86; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 8.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 258 TWHYYLNGRTATR 271
|||||

RESULT 20

ADY27058
ID ADY27058 standard; protein; 508 AA.

XX AC ADY27058;

XX DT 05-MAY-2005 (first entry)

XX DE Human inactive heparanase protein.

XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW protease; enzyme; enzyme purification.

XX OS Homo sapiens.

XX PN WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX PT Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.

XX PS Claim 257; SEQ ID NO 34; 211pp; English.

XX CC The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a human inactive heparanase protein used in the scope of the
CC invention.

XX SQ Sequence 508 AA;

Query Match 100.0%; Score 86; DB 9; Length 508;
Best Local Similarity 100.0%; Pred. No. 8.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
|||||

Db 259 TWHYYLNGRTATR 272

RESULT 21
ADZ19006
ID ADZ19006 standard; protein; 526 AA.
XX
AC ADZ19006;
XX
DT 16-JUN-2005 (first entry)
XX
DE HephYaluro construct protein.
XX
KW Enzyme engineering; heparanase; hephYaluro; metastasis;
KW autoimmune disease; inflammation; neoplasm; immune disorder;
KW antiinflammatory; cytostatic; immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
DR WPI: 2005-273382/28.
DR N-PSDB; ADZ19007.
XX
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
PS Example 2; SEQ ID NO 27; 65pp; English.
XX
SS The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hephYaluro construct
CC protein used in the scope of the invention.
XX
SQ Sequence 526 AA;
Query Match 100.0%; Score 86; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TWHYYLNGRTATR 14
|||||
Db 277 TWHYYLNGRTATR 290
RESULT 22
ABB07815
ID ABB07815 standard; protein; 527 AA.
XX
AC ABB07815;
XX
DT 03-JUL-2002 (first entry)
XX
DE Chicken signal peptide/human heparanase chimeric protein sequence.
XX
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.
Synthetic.
OS Gallus gallus.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "chicken heparanase signal peptide"
FT /note= "human heparanase mature protein"
XX
PN US2002034810-A1.
XX
PD 21-MAR-2002.
XX
PF 16-AUG-2001; 2001US-00930218.
XX
PR 20-SEP-2000; 2000US-00666390.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX
DR WPI: 2002-338926/37.
DR N-PSDB; ABL40753.
XX
PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
PT to treat various heparin-related disorders and the signal peptide is
PT useful in production of membrane-targeted or secreted recombinant
PT proteins.
XX
PS Disclosure; Page 26-28; 39pp; English.
XX
CC The invention relates to an isolated avian and reptile nucleic acid,
CC encoding a polypeptide with heparanase catalytic activity. The signal
CC peptide of the nucleic acid can be used to express membrane-associated or
CC secreted proteins in heterologous expression systems. The encoded
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC invasion, and to intervene with pathologies associated with impaired
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoa and bacterial infections or
CC disintegration of neurodegenerative plaques. The present sequence
CC represents a chicken signal peptide/human heparanase chimeric protein
CC sequence
XX
SQ Sequence 527 AA;
Query Match 100.0%; Score 86; DB 5; Length 527;
Best Local Similarity 100.0%; Pred. No. 8.5e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TWHYYLNGRTATR 14
|||||
Db 278 TWHYYLNGRTATR 291
RESULT 23
ABW02018
ID ABW02018 standard; protein; 527 AA.
XX
AC ABW02018;
XX
DT 12-FEB-2004 (first entry)
XX
DE Chimeric human-chicken heparanase protein.
XX
KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
KW wound healing; angiogenesis; restenosis; Genstmann-Strausler Syndrome;
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
KW immunosuppressive; vulnerary; bactericide; anti-angiogenic; virucide;

KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.
 XX
 OS Chimeric - Gallus gallus.
 OS Chimeric - Homo sapiens.
 XX
 PN US2003180788-A1.
 XX
 XX 25-SEP-2003.
 PD
 XX 08-MAY-2003; 2003US-00431438.
 XX
 XX 20-SEP-2000; 2000US-00666390.
 PR
 PR 16-AUG-2001; 2001US-00930218.
 XX
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;
 PI
 XX WPI; 2003-843931/78.
 DR
 DR N-PSDB; AAD63532.
 XX
 XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.
 PT
 XX Example; Page 26-28; Opp; English.
 PS
 XX The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genetmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
 CC some bacterial and protozoa infections. Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamine. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein
 XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 86; DB 7; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TWHYYLNGRTATR 14
 Db 278 TWHYYLNGRTATR 291
 RESULT 24
 ADO63825
 ID ADO63825 standard; protein; 527 AA.
 XX
 AC ADO63825;
 XX
 XX 26-AUG-2004 (first entry)
 DT
 XX Chimeric heparanase mutant E225A, SEQ ID:10.
 DE
 XX Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnuary; mutant; mutein.
 XX
 OS Homo sapiens.

OS Gallus gallus.
 OS Synthetic.
 OS Chimeric.
 XX
 FH Key
 FT Peptide
 FT
 FT Region
 FT
 FT Misc-difference 209
 FT /note= "Chicken heparanase signal peptide"
 FT 1..18
 FT 19..527
 FT /note= "Corresponds to residues 35-543 of human
 FT heparanase mutant E225A (SEQ ID NO:7)"
 FT 209
 FT /note= "Ala replaces wild-type Glu (active site proton
 FT donor). Corresponds to residue 225 of human heparanase
 FT mutant E225A (SEQ ID NO:7)"
 FT 327
 FT /note= "Active site nucleophile"
 XX
 PN WO2004048558-A2.
 XX
 XX 10-JUN-2004.
 PD
 XX 24-NOV-2003; 2003WO-IL000989.
 PF
 XX 24-NOV-2002; 2002IL-00153059.
 PR
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA
 XX Vlodaysky I, Zcharia E, Goldshmidt O, Ilan N;
 PI
 XX WPI; 2004-450373/42.
 DR
 DR N-PSDB; ADO63819.
 XX
 XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 PT
 XX Claim 10; SEQ ID NO 10; 128pp; English.
 PS
 XX The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may (ADO63825-
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase mutant E225A.
 XX
 SQ Sequence 527 AA;

Query Match 100.0%; Score 86; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14

Db	278	TWHYYLNGRTATR 291	
RESULT 25			
AD063827			
ID	AD063827	standard; protein; 527 AA.	
XX	AD063827;		
AC			
DT	26-AUG-2004	(first entry)	
XX			
DE	Chimeric heparanase mutant E225A/E343A, SEQ ID:12.		
XX			
KW	Human; chicken; heparanase; heparanase-derived protein; heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion; tissue sealant; injury; blood loss; endothelialisation; blood vessel; vascular graft; platelet adhesion; platelet aggregation; adhesion disorder; LAD; leukocyte adhesion deficiency; platelet aggregation; Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening; vulnerary; mutant; mutein.		
OS	Homo sapiens.		
OS	Gallus gallus.		
OS	Synthetic.		
OS	Chimeric.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1. .18	
FT	Region	/note= "Chicken heparanase signal peptide"	
FT		19. .527	
FT		/note= "Corresponds to residues 35-543 of human heparanase mutant E225A/E343A (SEQ ID NO:9)"	
FT	Misc-difference	209	
FT		/note= "Ala replaces wild-type Glu (active site proton donor). Corresponds to residue 225 of human heparanase mutant E225A/E343A (SEQ ID NO:9)"	
FT	Misc-difference	327	
FT		/note= "Ala replaces wild-type Glu (active site nucleophile). Corresponds to residue 343 of human heparanase mutant E225A/E343A (SEQ ID NO:9)"	
XX			
PN	WO2004048558-A2.		
XX			
PD	10-JUN-2004.		
XX			
PF	24-NOV-2003; 2003WO-IL000989.		
XX			
PR	24-NOV-2002; 2002IL-00153059.		
XX			
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.		
XX			
PI	Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;		
XX			
DR	WPI; 2004-450373/42.		
DR	N-PSDB; AD063821.		
XX			
PT	New nucleic acid construct comprising heparanase-derived polypeptide, useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's thrombasthenia, or Bernard-Soulier syndrome.		
XX			
PS	Claim 10; SEQ ID NO 12; 128pp; English.		
XX			
CC	The invention relates to nucleic acid constructs comprising a nucleic acid encoding a heparanase-derived protein which lacks heparanase endoglycosidase catalytic activity but which retains its cell-cell and cell-matrix adhesion properties. The constructs of the invention CC optionally further comprise operably linked regulatory elements. The CC invention also relates to the heparanase-derived proteins and host cells comprising the nucleic acid constructs of the invention. The heparanase- CC derived proteins are especially mutants of human heparanase in which the CC active site proton donor Glu225 and/or the active site nucleophile Glu343 CC are replaced with Ala (AD063822-AD063824), and the proteins may		
CC	optionally further comprise an avian heparanase signal peptide (AD063825-AD063827). The heparanase-derived protein, nucleic acid construct and host cells are useful in preparing a tissue sealant composition for sealing injuries, reducing the loss of blood, accelerating the healing and homeostasis of an injury, accelerating blood vessel endothelium formation or the endothelialisation of vascular grafts, accelerating the adhesive activity of mammalian cells, and accelerating the adhesion and aggregation of platelets. They may also be used in the treatment of disorders associated with adhesion deficiency such as LAD (leukocyte adhesion deficiency), Glanzmann's thrombasthenia (defective platelet function), or Bernard-Soulier syndrome (deficient platelet adhesion). The cells of the invention may additionally be used to screen for modulators of cell-cell and cell-matrix adhesion, and to prepare an implantable synthetic vascular graft comprising a tube made of a biocompatible material lined with the cells. The present sequence represents a chimeric protein comprising the signal peptide of chicken heparanase and residues 35-543 of the human heparanase double mutant E225A/E343A.		
XX			
SQ	Sequence 527 AA;		
	Query Match	100.0%; Score 86; DB 8; Length 527;	
	Best Local Similarity	100.0%; Pred. No. 8.5e-06;	
	Matches	14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	TWHYYLNGRTATR 14	
Db	278	TWHYYLNGRTATR 291	
RESULT 26			
AD063826			
ID	AD063826	standard; protein; 527 AA.	
XX			
AC	AD063826;		
XX			
DT	26-AUG-2004	(first entry)	
XX			
DE	Chimeric heparanase mutant E343A, SEQ ID:11.		
XX			
KW	Human; chicken; heparanase; heparanase-derived protein; heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion; tissue sealant; injury; blood loss; endothelialisation; blood vessel; vascular graft; platelet adhesion; platelet aggregation; adhesion disorder; LAD; leukocyte adhesion deficiency; Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening; vulnerary; mutant; mutein.		
OS	Homo sapiens.		
OS	Gallus gallus.		
OS	Synthetic.		
OS	Chimeric.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1. .18	
FT	Region	/note= "Chicken heparanase signal peptide"	
FT		19. .527	
FT		/note= "Corresponds to residues 35-543 of human heparanase mutant E343A (SEQ ID NO:8)"	
FT	Active-site	209	
FT		/note= "Active site proton donor"	
FT	Misc-difference	327	
FT		/note= "Ala replaces wild-type Glu (active site nucleophile). Corresponds to residue 343 of human heparanase mutant E343A (SEQ ID NO:8)"	
XX			
PN	WO2004048558-A2.		
XX			
PD	10-JUN-2004.		
XX			
PF	24-NOV-2003; 2003WO-IL000989.		
XX			
PR	24-NOV-2002; 2002IL-00153059.		
XX			
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.		
XX			
PI	Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;		
XX			
DR	WPI; 2004-450373/42.		
DR	N-PSDB; AD063821.		
XX			
PT	New nucleic acid construct comprising heparanase-derived polypeptide, useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's thrombasthenia, or Bernard-Soulier syndrome.		
XX			
PS	Claim 10; SEQ ID NO 12; 128pp; English.		
XX			
CC	The invention relates to nucleic acid constructs comprising a nucleic acid encoding a heparanase-derived protein which lacks heparanase endoglycosidase catalytic activity but which retains its cell-cell and cell-matrix adhesion properties. The constructs of the invention CC optionally further comprise operably linked regulatory elements. The CC invention also relates to the heparanase-derived proteins and host cells comprising the nucleic acid constructs of the invention. The heparanase- CC derived proteins are especially mutants of human heparanase in which the CC active site proton donor Glu225 and/or the active site nucleophile Glu343 CC are replaced with Ala (AD063822-AD063824), and the proteins may		

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
 PI
 XX
 DR WPI; 2004-450373/42.
 DR N-PSDB; ADO63820.
 XX
 XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 XX
 PS Claim 10; SEQ ID NO 11; 128pp; English.
 XX
 CC The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase mutant E343A.
 XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 86; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TWHHYLYNGRTATR 14
 |||||
 Db 278 TWHHYLYNGRTATR 291
 RESULT 27
 ADZ19004
 ID ADZ19004 standard; protein; 527 AA.
 AC
 XX
 AC ADZ19004;
 XX
 DT 16-JUN-2005 (first entry)
 DE HepGS4 construct protein.
 XX
 KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX
 XX Lahm A, Nardella C, Pallao M, Steinkuhler C;
 PI
 XX WPI; 2005-273382/28.
 DR N-PSDB; ADZ19002.
 XX
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Example 2; SEQ ID NO 25; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepGS4 construct protein
 CC used in the scope of the invention.
 XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 86; DB 9; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 TWHHYLYNGRTATR 14
 |||||
 Db 278 TWHHYLYNGRTATR 291
 RESULT 28
 AAY17083
 ID AAY17083 standard; protein; 532 AA.
 XX
 AC AAY17083;
 XX
 DT 21-JUL-1999 (first entry)
 XX
 DE Seq ID No: 15 of WO9921975.
 XX
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
 KW human; HSPG.
 XX
 OS Homo sapiens.
 XX
 PN WO9921975-A1.
 XX
 PD 06-MAY-1999.
 XX
 XX 28-OCT-1998; 98WO-AU000898.
 PF
 XX 28-OCT-1997; 97AU-0000062.
 PR
 PR 09-DEC-1997; 97AU-00000812.
 XX
 XX (AUSU) UNIV AUSTRALIAN NAT.
 FA
 XX Freeman CG, Hulett MD, Parish CR, Handorf BJ;
 PI
 XX WPI; 1999-312956/26.
 DR N-PSDB; AAX37260.
 XX
 PT Polynucleotides encoding mammalian endoglucuronidases, especially
 PT heparanases, useful to promote wound healing.

XX Claim 6; Page 76-79; 112pp; English.

PS The invention relates to nucleic acid sequences that encode heparanase

CC enzymes having endoglucuronidase activity. Recombinant heparanases are

CC capable of removing the HS side chain from heparan sulfate proteoglycan

CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to

CC inhibit heparanase, this is useful for treatment of a physiological or

CC medical condition associated with elevated heparanase activity, such as

CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,

CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially

CC associated with tissue development and repair. The conditions mentioned

CC above can be diagnosed using specific antibodies, and also using primers

CC and probes specific for the heparanase polynucleotides. Other uses of the

CC heparanases include sequencing sulfated molecules such as HSPG

XX

SQ Sequence 532 AA;

Query Match 100.0%; Score 86; DB 2; Length 532;

Best Local Similarity 100.0%; Pred. No. 8.6e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLNGRTATR 14

DB 294 TWHHYLNGRTATR 307

|||||

RESULT 29

AA02345

ID AAY02345 standard; protein; 543 AA.

XX

AC AAY02345;

XX

DT 09-JUL-1999 (first entry)

XX

DE A human heparanase protein.

XX

XX Heparanase; hp; modulator; heparin-binding growth factor;

KW cellular response; cytokine; cell interaction; plasma lipoprotein;

KW cellular susceptibility; infection; disintegration;

KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;

KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;

KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX

OS Homo sapiens.

XX

PN WO9911798-A1.

XX

PD 11-MAR-1999.

XX

PF 31-AUG-1998; 98WO-US017954.

XX

PR 02-SEP-1997; 97US-00922170.

PR 02-JUL-1998; 98US-00109386.

XX

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (FRIE/) FRIEDMAN M M.

XX

XX Pecker I, Vlodavsky I, Feinstein E;

XX

DR WPI; 1999-302255/25.

DR N-PSDB; AAX35648.

XX

XX New human polynucleotide useful for treating angiogenesis, restenosis,

PT and inflammation.

XX

PS Claim 6; Fig 1; 63pp; English.

XX

XX The specification describes a polypeptide having heparanase (hp)

CC activity. The recombinant protein is used as a modulator of heparin-

CC binding growth factors, cellular responses to heparin-binding growth

CC factors and cytokines, cell interaction with plasma lipoproteins,

CC cellular susceptibility to viral, protozoal and bacterial infections or

CC disinfection of neurodegenerative plaques. Heparanase may be useful for

CC conditions such as wound healing, angiogenesis, restenosis,

CC atherosclerosis, inflammation, neurodegenerative diseases, and viral

CC infections. Mammalian heparanase can be used to neutralize plasma

CC heparin, and anti-heparanase antibodies may be applied for

CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and

CC renal failure in biopsy specimens, plasma samples, and body fluids. The

XX present sequence represents human heparanase

SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 2; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLNGRTATR 14

DB 294 TWHHYLNGRTATR 307

|||||

RESULT 30

AA017082

ID AAY17082 standard; protein; 543 AA.

XX

AC AAY17082;

XX

DT 21-JUL-1999 (first entry)

XX

DE Human heparanase enzyme.

XX

KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;

KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;

KW arteriosclerosis; atherosclerosis; inflammation; tissue development;

KW human; HSPG.

XX

OS Homo sapiens.

XX

PN WO9921975-A1.

XX

PD 06-MAY-1999.

XX

PF 28-OCT-1998; 98WO-AU000898.

XX

PR 28-OCT-1997; 97AU-00000062.

PR 09-DEC-1997; 97AU-00000812.

XX

PA (AUSU) UNIV AUSTRALIAN NAT.

XX

PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

XX

DR WPI; 1999-312956/26.

DR N-PSDB; AAX37259.

XX

PT Polynucleotides encoding mammalian endoglucuronidases, especially

XX heparanases, useful to promote wound healing.

XX

PS Claim 6; Page 69-73; 112pp; English.

XX

CC The invention relates to nucleic acid sequences that encode heparanase

CC enzymes having endoglucuronidase activity. Recombinant heparanases are

CC capable of removing the HS side chain from heparan sulfate proteoglycan

CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to

CC inhibit heparanase, this is useful for treatment of a physiological or

CC medical condition associated with elevated heparanase activity, such as

CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,

CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially

CC associated with tissue development and repair. The conditions mentioned

CC above can be diagnosed using specific antibodies, and also using primers

CC and probes specific for the heparanase polynucleotides. Other uses of the

CC heparanases include sequencing sulfated molecules such as HSPG. The

CC present sequence represents a human heparanase
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
|||||
Db 294 TWHYYLNGRTATR 307

RESULT 31
AA57590
ID AAY57590 standard; protein; 543 AA.
XX
AC AAY57590;
XX
DT 02-MAR-2000 (first entry)
XX
DE Human heparanase.
XX
KW Human; heparanase; hpa; genetic modification; expression; anticancer;
KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW micrometastasis; autoimmune lesion; kidney failure.
XX
OS Homo sapiens.
XX
PN WO9957244-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US009256.
XX
PR 01-MAY-1998; 98US-00071618.
XX
PR 02-MAR-1999; 99US-00260038.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (FRIE/) FRIEDMAN M M.
XX
PI Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;
PI Peleg Y, Shlomi Y;
XX
WPI; 2000-062144/05.
DR N-PSDB; AA239195.
XX
PT Engineered cells that express recombinant heparanase, useful
PT therapeutically, e.g. for treating angiogenesis and to screen for
PT specific inhibitors, potential anticancer agents.
XX
PS Claim 3; Page 107-109; 118pp; English.
XX
CC The present invention describes genetically modified cells (A) containing
CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
CC and express recombinant heparanase (II). Heparanase cleaves heparan
CC sulphate (HS) at specific intrachain sites, resulting in release of
CC heparin-binding growth factors, enzymes and proteins that are sequestered
CC by HS in basement membranes, extracellular matrix or cell surfaces. It
CC may also be implicated in tumour angiogenesis and metastases. (II) is
CC potentially useful in wound healing and for treating angiogenesis,
CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
CC infection and cystic fibrosis. It can also be used to neutralise heparin
CC (an alternative to protamine) and to screen for specific inhibitors
CC (potentially useful for treating cancer and metastases). Antibodies
CC raised against (II) are used for immunodetection and diagnosis of
CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
CC in large quantities, in a form that is homogeneously processed and
CC activated/neutralised by a dedicated protease. The present sequence

CC represents human heparanase
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
|||||
Db 294 TWHYYLNGRTATR 307

RESULT 32
AAB08849
ID AAB08849 standard; protein; 543 AA.
XX
AC AAB08849;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of a human heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.
XX
OS Homo sapiens.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US003542.
XX
PR 01-MAR-1999; 99US-00258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
WPI; 2000-579289/54.
DR N-PSDB; AAA75051.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
XX
PS Claim 22; Fig 1; 152pp; English.
XX
CC The present sequence represents a human protein with heparanase catalytic
CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC particularly in treating tumour, inflammation or autoimmunity.
CC Particularly, the polynucleotide is useful in modulating the
CC bioavailability of heparin-binding growth factors, cellular responses to
CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
CC susceptibility to certain viral and some bacterial and protozoa
CC infections, or disintegration of neurodegenerative plaques. The
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
CC radiation burns), and in the treatment of angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC bacterial or protozoa infections
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;

```

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 33
AA52990
ID AAY52990 standard; protein; 543 AA.
XX AAY52990;
XX
XX 21-FEB-2000 (first entry)
XX Human heparanase protein sequence.
XX
XX Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
XX antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
XX metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
XX mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
XX inflammation; haemorrhagic nephritis; nephrotic syndrome;
XX autoimmune disease; anticancer; kidney disease.
XX
XX Homo sapiens.
XX
XX W09957153-A1.
XX
XX 11-NOV-1999.
XX
XX 29-APR-1999; 99WO-US009255.
XX
XX 01-MAY-1998; 98US-00071739.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX WPI; 2000-052944/04.
XX N-PSDB; AA233290.
XX
XX Heparanase-specific molecular probes useful for diagnosis and treatment,
XX e.g. of tumors, and for targeted drug delivery.
XX
XX Example; Page 81-82; 90pp; English.
XX
XX The present invention describes heparanase-specific molecular probes,
XX useful for methods of detecting heparanase in situ. The probes and anti-
XX heparanase antibodies are used to detect or quantify the expression of
XX heparanase, for diagnosis and monitoring of diseases (especially
XX metastasis), for treatment of heparanase-associated diseases (e.g.
XX tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
XX mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
XX metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
XX colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
XX diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
XX sepsis and inflammatory or autoimmune disease), for targeted drug
XX delivery (e.g. of anticancer agents) and as research reagents. The
XX present sequence represents human heparanase, which is used in the
XX exemplification of the present invention
XX
XX Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 34
AA52990
ID AAY52990 standard; protein; 543 AA.
XX AAY52990;
XX
XX 20-APR-2001 (first entry)
XX Human heparanase protein sequence.
XX
XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
XX atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
XX neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
XX gene therapy; human.
XX
XX Homo sapiens.
XX
XX W0200100643-A2.
XX
XX 04-JAN-2001.
XX
XX 19-JUN-2000; 2000WO-IL000358.
XX
XX 25-JUN-1999; 99US-0140801P.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
XX Pecker I, Michal I, Itzhaki H;
XX WPI; 2001-137930/14.
XX
XX New polynucleotides and polypeptides that are distantly homologous to
XX heparanase, useful in wound healing, as well as in gene therapy protocols
XX for angiogenesis, restenosis, atherosclerosis, or inflammation.
XX
XX Disclosure; Page 64-65; 67pp; English.
XX
XX This sequence represents a heparanase of the invention. The heparanase
XX DNA and protein sequences are useful in wound healing, angiogenesis,
XX restenosis, atherosclerosis, inflammation, pulmonary diseases,
XX neurodegenerative diseases (such as Scrape, Alzheimer's disease, and
XX Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
XX sequence is particularly useful in gene therapy
XX
XX Sequence 543 AA;

Query Match 100.0%; Score 86; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 35
AA86206
ID AAB86206 standard; protein; 543 AA.
XX AAB86206;
XX
XX 24-AUG-2001 (first entry)
XX
XX Human heparanase inhibitor protein.
XX
XX Heparanase; inhibitor; cardiac insufficiency; cardiast; nephrotropic;
XX hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
XX primary cardiomyopathy; peripheral odema; pulmonary congestion;
XX hepatic congestion; hydrothorax; ascite; nocturia; human.
XX
XX Homo sapiens.
XX

```

PN DE19955803-A1.
XX 23-MAY-2001.
XX 19-NOV-1999; 99DE-01055803.
XX 19-NOV-1999; 99DE-01055803.
XX (KNOL) KNOLL AG.
XX Herr D, Hahn A, Laux V;
XX WPI: 2001-368371/39.
XX N-PSDB; AAF20940.
PT Treatment or prevention of cardiac insufficiency and related conditions,
PT e.g. pulmonary congestion and dyspnoea, comprises administration of
PT heparanase inhibitor.
XX
XX Disclosure; Page 11-13; 16pp; German.
XX This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
CC the invention has cardiac, nephrotropic and hepatotropic activity. The
CC products of the invention can be used in human and veterinary medicine,
CC for the treatment or prevention of congestive heart failure e.g. primary
CC cardiomyopathy. Associated conditions treated or prevented with the
CC inhibitor are especially peripheral odemas, pulmonary and hepatic
CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
CC nocturia can also be treated. This sequence represents the human
CC heparanase protein described in the method of the invention
XX
XX Sequence 543 AA;
Query Match 100.0%; Score 86; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TWHYYLNGRTATR 14
||| ||||| ||||| |||||
Db 294 TWHYYLNGRTATR 307
RESULT 36
AAB88361
ID AAB88361 standard; protein; 543 AA.
XX
XX AAB88361;
XX 23-MAY-2001 (first entry)
XX Human membrane or secretory protein clone PSEC0090.
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
XX rheumatoid arthritis; diabetes.
XX Homo sapiens.
XX
XX EP1067182-A2.
XX 10-JAN-2001.
XX
XX 07-JUL-2000; 2000EP-00114090.
XX
XX 08-JUL-1999; 99JP-00194179.
XX 11-JAN-2000; 2000JP-00118775.
XX 02-MAY-2000; 2000JP-00183766.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX

DR WPI: 2001-093989/11.
XX N-PSDB; AAF93788.
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX gene therapy or as candidate target molecules in drug development.
XX
XX Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.
XX
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX which encode human secretory or membrane proteins represented by AAB88317
XX - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
XX AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
XX invention. The invention also includes methods for the production of
XX antibodies directed against the proteins, and cDNA sequences, which can
XX be used in vaccines. The polynucleotide sequences can be used in gene
XX therapy. The polynucleotide sequences and the proteins they encode may be
XX used in the prevention, treatment and diagnosis of diseases associated
XX with inappropriate secretory protein/membrane protein expression. The
XX nucleic acids and complementary sequences may also be used as DNA probes
XX in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
XX and quantitate the presence of similar nucleic acid sequences in samples.
XX They may also be used to study the expression and function of secretory
XX proteins/membrane polypeptides and their role in metabolism. The
XX polypeptides may be used as antigens in the production of antibodies
XX against them and in assays to identify modulators (agonists and
XX antagonists) of expression and activity. The antibodies and antagonists
XX may also be used as therapeutic agents to down regulate expression and
XX activity. The antibodies may also be used as diagnostic agents for
XX detecting the presence of the polypeptides in samples (e.g. by enzyme
XX linked immunosorbant assay (ELISA)). Examples of diseases which may be
XX treated include rheumatoid arthritis and diabetes
XX
XX Sequence 543 AA;
Query Match 100.0%; Score 86; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TWHYYLNGRTATR 14
||| ||||| ||||| |||||
Db 294 TWHYYLNGRTATR 307
RESULT 37
AAB07813
ID ABB07813 standard; protein; 543 AA.
XX
XX ABB07813;
XX 03-JUL-2002 (first entry)
XX Human heparanase sequence.
XX
XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
XX anti-protozoan; neuroprotective; heparin; human.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..35
XX /note= "signal peptide"
XX Protein 36..543
XX /note= "mature protein"
XX
XX US2002034810-A1.
XX 21-MAR-2002.
XX
XX 16-AUG-2001; 2001US-00930218.
XX 20-SEP-2000; 2000US-00666390.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
 XX WPI; 2002-338926/37.
 XX
 PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.
 XX
 XX Disclosure; Fig 1a; 39pp; English.
 XX
 CC The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a human heparanase protein sequence used in similarity studies
 XX
 XX Sequence 543 AA;
 SQ
 Query Match 100.0%; Score 86; DB 5; Length 543;
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 Db |||||
 294 TWHYYLNGRTATR 307
 RESULT 38
 ADD18950
 ID ADD18950 standard; protein; 543 AA.
 XX
 AC ADD18950;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human disease related protein SeqID439.
 XX
 KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnery; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transportation; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.
 XX
 OS Homo sapiens.
 XX
 XX WO2003018621-A2.
 PN
 XX 06-MAR-2003.
 PD
 XX 23-AUG-2002; 2002WO-GB003892.
 PF
 XX 23-AUG-2001; 2001GB-00020558.
 PR
 XX 05-OCT-2001; 2001GB-00024037.
 XX
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA
 XX
 XX Kingman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
 PI WPI; 2003-290046/28.
 DR
 DR N-PSDB; ADD18951.
 XX
 XX New substantially purified polypeptide, useful for diagnosing or treating

PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.
 XX
 XX Claim 25; SEQ ID NO 439; 424pp; English.
 XX
 CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 86; DB 7; Length 543;
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TWHYYLNGRTATR 14
 Db |||||
 294 TWHYYLNGRTATR 307
 RESULT 39
 ADG88800
 ID ADG88800 standard; protein; 543 AA.
 XX
 AC ADG88800;
 XX
 DT 11-MAR-2004 (first entry)
 XX
 XX Human hpa protein.
 XX
 KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
 KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.
 XX
 OS Homo sapiens.
 XX
 XX US2003161823-A1.
 PN
 XX 28-AUG-2003.
 PD
 XX 14-JAN-2003; 2003US-00341582.
 PF
 XX 31-AUG-1998; 98WO-US017954.
 PR
 XX 01-MAR-1999; 99US-00258892.
 PR
 XX 06-FEB-2001; 2001US-00776874.
 PR
 XX 05-SEP-2001; 2001WO-IL000830.
 PR
 XX 19-NOV-2001; 2001US-00988113.
 XX
 XX (ILAN/) ILAN N.
 PA (VLOD/) VLODAVSKY I.
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PECK/) PECKER I.
 PA (FEIN/) FEINSTEIN E.
 XX
 XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
 PI WPI; 2003-897910/82.
 DR
 DR N-PSDB; ADG88799, ADG88801, ADG88832.
 XX
 XX Composition for treating a wound comprising recombinant heparanase is
 XX useful to induce or accelerate wound healing and induce or accelerate
 PT angiogenesis.

XX PS Claim 2; SEQ ID NO 10; 143pp; English.

XX CC The present invention relates to methods and compositions for inducing

XX CC and/or accelerating wound healing via the catalytic activity of

XX CC heparanase. The invention is used to induce or accelerate a healing

XX CC process, particularly of an ulcer, burn, laceration, surgical incision,

XX CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate

XX CC angiogenesis. The present sequence is human hpa protein.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 7; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14

Db 294 TWHYYLNGRTATR 307

|||||

RESULT 40

ADL16379

ID ADL16379 standard; protein; 543 AA.

XX AC ADL16379;

XX DT 06-MAY-2004 (first entry)

XX DE Human heparanase partial protein.

XX KW Human; heparanase; heparanase-dependent cancer; cancer;

XX KW autoimmune reaction; inflammation; chromosome 4; enzyme.

XX OS Homo sapiens.

XX US2003236215-A1.

XX 25-DEC-2003.

XX 09-JUN-2003; 2003US-00456573.

XX 31-AUG-1998; 98WO-US017954.

XX 01-MAR-1999; 99US-00258892.

XX 08-NOV-1999; 99US-00435739.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Pecker I, Vlodavsky I, Feinstein E;

XX WPI; 2004-070610/07.

XX New antisense oligonucleotide hybridizable with a polynucleotide encoding

XX a polypeptide with heparanase activity, useful for treating diseases such

XX as cancer and autoimmune disorders.

XX Claim 3; SEQ ID NO 10; 108pp; English.

XX The invention relates to an antisense oligonucleotide (ASO) comprising a

XX polynucleotide or a polynucleotide analogue of at least 10 bases being

XX hybridisable in vivo, under physiological conditions, with a portion of

XX a polynucleotide strand encoding a polypeptide having heparanase

XX catalytic activity. Also included are a method of in vivo downregulating

XX heparanase activity (comprising administering the ASO in vivo), a method

XX of treating a subject suffering from a pathological condition

XX (characterised by heparanase activity, comprising administering ASO to

XX the subject), a pharmaceutical composition comprising the ASO and a

XX carrier, an antisense nucleic acid construct (comprising a promoter

XX sequence and a polynucleotide sequence directing the synthesis of an

XX antisense RNA sequence of at least 10 bases being hybridisable in vivo,

XX under physiological conditions, with a polynucleotide strand encoding a

XX polypeptide having heparanase catalytic activity), a method of in vivo

CC downregulating heparanase activity (comprising administering in vivo the

CC antisense nucleic acid construct), a pharmaceutical composition

CC comprising the antisense nucleic acid construct and a carrier, and an

CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide

CC analogue of at least 10 bases being hybridisable in vivo, under

CC physiological conditions, with a portion of a polynucleotide strand being

CC characterised by forming at least a portion of an untranslated region

CC (UTR) for a polynucleotide strand encoding a polypeptide having

CC heparanase catalytic activity. The methods and compositions of the

CC present invention are useful for the prevention and/or treatment of

CC diseases or conditions associated with aberrant heparanase activity, such

CC as heparanase-dependent cancer, cancer, autoimmune reaction and

CC inflammation. The gene for human heparanase is located on chromosome 4.

XX The present sequence is a human heparanase protein.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14

Db 294 TWHYYLNGRTATR 307

|||||

RESULT 41

ADK52086

ID ADK52086 standard; protein; 543 AA.

XX AC ADK52086;

XX DT 20-MAY-2004 (first entry)

XX DE Human atopic dermatitis/psoriasis-associated protein #1.

XX KW Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;

XX KW antipsoriatic; rash.

XX OS Homo sapiens.

XX WO2004016785-A1.

XX 26-FEB-2004.

XX 06-AUG-2003; 2003WO-JP009999.

XX 06-AUG-2002; 2002JP-00229319.

XX 14-MAY-2003; 2003JP-00136544.

XX (GENO-) GENOX RES INC.

XX (UYJU-) UNIV JUNTENDO.

XX Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;

XX Mitsuishi K;

XX WPI; 2004-214514/20.

XX N-PSDB; ADK51968.

XX Detecting atopic dermatitis or psoriasis comprises assaying levels of

XX expression of an indicator gene at a rash site and non-rash site of a

XX person with atopic dermatitis or psoriasis.

XX Example 2; SEQ ID NO 119; 484pp; Japanese.

XX The invention relates to detecting atopic dermatitis or psoriasis

XX comprising assaying the levels of expression of an indicator gene at a

XX rash site and non-rash site of a person with atopic dermatitis or

XX psoriasis, comparing these levels with those of a healthy person, and

XX determining that if the levels of indicators are higher or lower, then

XX this indicates the disease. Also included are a reagent for detecting

XX atopic dermatitis or psoriasis, a kit for screening for treatments, a

XX transgenic non human vertebrate animal models for the diseases, an agent

CC for inducing the diseases in mice and a DNA chip for assaying for the
CC indicator genes. The method is used for treatment, detection and animal
CC models for research of atopic dermatitis and psoriasis. The present
CC sequence is a protein encoded by an indicator gene of the invention.
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
|||||
Db 294 TWHYYLNGRTATR 307

RESULT 42
ADM48716
ID ADM48716 standard; protein; 543 AA.
XX
AC ADM48716;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human hpa protein #1.
XX
KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW human.
XX
OS Homo sapiens.
XX
PN US2003217375-A1.
XX
PD 20-NOV-2003.
XX
PF 24-FEB-2003; 2003US-00371218.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 06-FEB-2001; 2001US-00776874.
PR 19-NOV-2001; 2001US-00988113.
XX
PA (ZCHA/) ZCHARIA E.
PA (VLOD/) VLODAVSKY I.
PA (METZ/) METZGER S.
PA (PECK/) PECKER I.
PA (ILAN/) ILAN N.
PA (CHAJ/) CHAJEK-SHAUL T.
PA (GOLD/) GOLDSHMIDT O.
XX
PI Zcharia E, Vlodayevsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX
WPI: 2004-021918/02.
DR N-PSDB; ADM48715, ADM48717.
XX
PT New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
PS Example 1; SEQ ID NO 10; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary

CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
|||||
Db 294 TWHYYLNGRTATR 307

RESULT 43
ADM48759
ID ADM48759 standard; protein; 543 AA.
XX
AC ADM48759;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human hpa protein #2.
XX
KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW human.
XX
OS Homo sapiens.
XX
PN US2003217375-A1.
XX
PD 20-NOV-2003.
XX
PF 24-FEB-2003; 2003US-00371218.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 06-FEB-2001; 2001US-00776874.
PR 19-NOV-2001; 2001US-00988113.
XX
PA (ZCHA/) ZCHARIA E.
PA (VLOD/) VLODAVSKY I.
PA (METZ/) METZGER S.
PA (PECK/) PECKER I.
PA (ILAN/) ILAN N.
PA (CHAJ/) CHAJEK-SHAUL T.
PA (GOLD/) GOLDSHMIDT O.
XX
PI Zcharia E, Vlodayevsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX
WPI: 2004-021918/02.
DR N-PSDB; ADM48748.
XX
PT New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
PS Example 10; Fig 16; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.

```

XX SQ Sequence 543 AA;
Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 44
ADN05074
ID ADN05074 standard; protein; 543 AA.
XX AC ADN05074;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #716.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX XX New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX XX Claim 9; SEQ ID NO 1468; 3069pp; English.
XX PS The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;

Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
Wu TD;
WPI; 2004-305105/28.
N-PSDB; ADN05073.
New PRO nucleic acid or polypeptide, useful for preparing a
pharmaceutical composition for diagnosing or treating psoriasis in a
mammal.
Claim 9; SEQ ID NO 1468; 3069pp; English.
The invention relates to novel polynucleotide and polypeptides for
treating psoriasis or a sequence having at least 80% identity to the
above sequences. The nucleic acid is useful for preparing a composition
for diagnosing or treating psoriasis in a mammal. This sequence
corresponds to one of the polypeptides of the invention.
Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 45
ADN04902
ID ADN04902 standard; protein; 543 AA.
XX AC ADN04902;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #631.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX XX New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX XX Claim 9; SEQ ID NO 1468; 3069pp; English.
XX PS The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 46
ADN063831
ID ADN063831 standard; protein; 543 AA.
XX AC ADN063831;
XX DT 26-AUG-2004 (first entry)
XX DE Human heparanase mutant E378A.
XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
XX KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
XX KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
XX KW vascular graft; platelet adhesion; platelet aggregation;
XX KW adhesion disorder; LAD; leukocyte adhesion deficiency;
XX KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
XX KW vulnery; mutant; mutein; enzyme.
XX OS Homo sapiens.
XX OS Synthetic.
XX XX Key Location/Qualifiers
XX FT Active-site 225 /note= "Active site proton donor"
XX FT Active-site 343 /note= "Active site nucleophile"
XX FT Misc-difference 378 /note= "Ala replaces wild-type Glu"

```

```

DE Antipsoriatic protein sequence #631.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX XX New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX XX Claim 9; SEQ ID NO 1296; 3069pp; English.
XX PS The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 46
ADN063831
ID ADN063831 standard; protein; 543 AA.
XX AC ADN063831;
XX DT 26-AUG-2004 (first entry)
XX DE Human heparanase mutant E378A.
XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
XX KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
XX KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
XX KW vascular graft; platelet adhesion; platelet aggregation;
XX KW adhesion disorder; LAD; leukocyte adhesion deficiency;
XX KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
XX KW vulnery; mutant; mutein; enzyme.
XX OS Homo sapiens.
XX OS Synthetic.
XX XX Key Location/Qualifiers
XX FT Active-site 225 /note= "Active site proton donor"
XX FT Active-site 343 /note= "Active site nucleophile"
XX FT Misc-difference 378 /note= "Ala replaces wild-type Glu"

```

XX WO2004048558-A2.
XX 10-JUN-2004.
XX 24-NOV-2003; 2003WO-IL000989.
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX Example 4; Page; 128pp; English.
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The
XX invention also relates to the heparanase-derived proteins and host cells
XX comprising the nucleic acid constructs of the invention. The heparanase-
XX derived proteins are especially mutants of human heparanase in which the
XX active site proton donor Glu225 and/or the active site nucleophile Glu343
XX are replaced with Ala (ADO63822-ADO63824), and the proteins may
XX optionally further comprise an avian heparanase signal peptide (ADO63825-
XX ADO63827). The heparanase-derived protein, nucleic acid construct and
XX host cells are useful in preparing a tissue sealant composition for
XX sealing injuries, reducing the loss of blood, accelerating the healing
XX and homeostasis of an injury, accelerating blood vessel endothelium
XX formation or the endothelialisation of vascular grafts, accelerating the
XX adhesive activity of mammalian cells, and accelerating the adhesion and
XX aggregation of platelets. They may also be used in the treatment of
XX disorders associated with adhesion deficiency such as LAD (leukocyte
XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
XX function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
XX cells of the invention may additionally be to screen for modulators of
XX cell-cell and cell-matrix adhesion, and to prepare an implantable
XX synthetic vascular graft comprising a tube made of a biocompatible
XX heparanase mutant E378A created in an example of the invention which
XX retains its heparanase catalytic activity. The present sequence is not
XX shown in the invention, but is derived from the protein sequence of
XX GenBank accession number AF144325 and the information provided on page
XX 70.
XX Sequence 543 AA;
XX Query Match 100.0%; Score 86; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 8.8e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307
RESULT 47
ID ADO63824
ID ADO63824 standard; protein; 543 AA.
XX ADO63824;
XX
XX 26-AUG-2004 (first entry)
XX Human heparanase mutant E225A/E343A, SEQ ID:9.
XX

KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnerary; mutant; mutein.
OS Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 225 /note= "Ala replaces wild-type Glu (active site proton
XX donor)"
XX FT
XX Misc-difference 343 /note= "Ala replaces wild-type Glu (active site
XX nucleophile)"
XX PN WO2004048558-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-IL000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX N-PSDB; ADO63818.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX Claim 9; SEQ ID NO 9; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The
XX invention also relates to the heparanase-derived proteins and host cells
XX comprising the nucleic acid constructs of the invention. The heparanase-
XX derived proteins are especially mutants of human heparanase in which the
XX active site proton donor Glu225 and/or the active site nucleophile Glu343
XX are replaced with Ala (ADO63822-ADO63824), and the proteins may
XX optionally further comprise an avian heparanase signal peptide (ADO63825-
XX ADO63827). The heparanase-derived protein, nucleic acid construct and
XX host cells are useful in preparing a tissue sealant composition for
XX sealing injuries, reducing the loss of blood, accelerating the healing
XX and homeostasis of an injury, accelerating blood vessel endothelium
XX formation or the endothelialisation of vascular grafts, accelerating the
XX adhesive activity of mammalian cells, and accelerating the adhesion and
XX aggregation of platelets. They may also be used in the treatment of
XX disorders associated with adhesion deficiency such as LAD (leukocyte
XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
XX function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
XX cells of the invention may additionally be to screen for modulators of
XX cell-cell and cell-matrix adhesion, and to prepare an implantable
XX synthetic vascular graft comprising a tube made of a biocompatible
XX heparanase mutant E378A created in an example of the invention which
XX retains its heparanase catalytic activity. The present sequence is not
XX shown in the invention, but is derived from the protein sequence of
XX GenBank accession number AF144325 and the information provided on page
XX 70.
XX Sequence 543 AA;
XX Query Match 100.0%; Score 86; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 8.8e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents the human
CC heparanase mutant E343A.

XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHHYILNGRTATR 14
| | | | | | | | | | | | | | | |
Db 294 TWHHYILNGRTATR 307

RESULT 49
ADO63832
ID ADO63832 standard; protein; 543 AA.
XX
AC ADO63832;
XX
XX
DT 26-AUG-2004 (first entry)
XX
DE Human heparanase mutant E396A.
XX
KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnerary; mutant; mucin; enzyme.
XX
OS Homo sapiens.
OS Synthetic.
OS
XX
XX
FH Key Location/Qualifiers
FT Active-site 225 /note= "Active site proton donor"
FT FT 343
FT Active-site /note= "Active site nucleophile"
FT FT 396
FT Misc-difference /note= "Ala replaces wild-type Glu"
XX
XX WO2004048558-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-IL000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX
XX WPI; 2004-450373/42.
XX
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Example 4; Page; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The

CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a human
 CC heparanase mutant E378A created in an example of the invention which
 CC retains its heparanase catalytic activity. The present sequence is not
 CC shown in the invention, but is derived from the protein sequence of
 CC GenBank accession number AF144325 and the information provided on page
 CC 70.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 8.8e-06; Indels 0; Gaps 0;
 Matches 14; Conservative 0; Mismatches 0;

Qy 1 TWHYYLNGRTATR 14
 |||||
 Db 294 TWHYYLNGRTATR 307

RESULT 50

ADO63822
 ID ADO63822 standard; protein; 543 AA.

XX AC ADO63822;

XX DT 26-AUG-2004 (first entry)

XX DE Human heparanase mutant E225A, SEQ ID:7.

XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein.

XX OS Homo sapiens.
 OS Synthetic.

XX FH Key Location/Qualifiers

FT Misc-difference 225
 /note= "Ala replaces wild-type Glu (active site proton
 FT donor)"
 FT Active-site 343
 FT /note= "Active site nucleophile"

XX PN WO2004048558-A2.

XX PD 10-JUN-2004.

XX PF 24-NOV-2003; 2003WO-IL000989.

XX PR 24-NOV-2002; 2002IL-00153059.

XX

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

XX DR WPI; 2004-450373/42.

XX DR N-PSDB; ADO63816.

XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.

XX PS Claim 9; SEQ ID NO 7; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents the human
 CC heparanase mutant E225A.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
 |||||
 Db 294 TWHYYLNGRTATR 307

Search completed: June 5, 2006, 12:42:27
 Job time : 121.425 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 15.5342 Seconds
(without alignments) 86.714 Million cell updates

Title: US-10-645-659A-10
Perfect score: 86
Sequence: 1 TWHYYVLNGRTATR

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

```
Minimum DB seq length: 0
Maximum DB seq length: 20
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1000 summaries

```
Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	51	59.3	1179	2	F71190	probable chromosome
2	46	53.5	350	2	J5251	beta-galactoside a
3	46	53.5	350	2	BC420	beta-galactoside a
4	45	52.3	190	2	T01245	N-acetyltransferas
5	44	51.2	290	2	AC2030	hypothetical prote
6	43	50.0	342	2	S55675	Gal-beta-1,3GalNac
7	43	50.0	479	2	D85330	hypothetical prote
8	43	50.0	480	2	JC7506	heparanase protein
9	43	50.0	481	2	T04607	hypothetical prote
10	43	50.0	1177	2	B75150	chromosome segrega
11	43	50.0	1200	2	S77524	chromosome segrega
12	42.5	49.4	1035	2	B90656	icmF-like protein
13	42.5	49.4	1144	2	C85507	probable macrophag
14	42	48.8	337	2	S36824	beta-galactoside a
15	42	48.8	340	2	I54229	beta-galactoside a
16	42	48.8	343	2	A45073	Gal beta 1,3GalNAC
17	42	48.8	396	2	AD3012	conserved hypothe
18	42	48.8	492	2	C98272	hypothetical prote
19	42	48.8	574	2	A46054	GTP-binding protei
20	42	48.8	1553	2	S67483	adenosinetriphosph
21	41	47.7	216	2	A61259	glycoprotein S - p
22	41	47.7	305	2	AB2349	polysaccharide dea
23	41	47.7	353	2	G82577	phage-related inte
24	41	47.7	463	2	T41390	zinc finger protei
25	41	47.7	473	2	E81182	hypothetical prote
26	41	47.7	473	2	B81919	probable membrane
27	41	47.7	1225	1	S24284	E2 glycoprotein pr
28	41	47.7	1225	2	S46607	E2 glycoprotein -
29	41	47.7	1447	1	VG1HE3	E2 glycoprotein pr

103	38	44.2	691	2	JE0150	acetylcholinesterase	176	36.5	42.4	192	2	S42999	viral infectivity
104	38	44.2	746	2	A25363	acetylcholinesterase	177	36.5	42.4	192	2	S42991	viral infectivity
105	37.5	43.6	192	2	B53294	superoxide dismutase	178	36.5	42.4	192	2	S42949	viral infectivity
106	37.5	43.6	199	2	T50046	superoxide dismutase	179	36.5	42.4	192	2	S42993	viral infectivity
107	37.5	43.6	200	2	T50045	superoxide dismutase	180	36.5	42.4	200	2	A34319	superoxide dismutase
108	37.5	43.6	546	2	JC4798	seizure-related gene	181	36.5	42.4	200	2	H84274	superoxide dismutase
109	37.5	43.6	550	2	I37579	DiGeorge syndrome	182	36.5	42.4	200	2	T50043	superoxide dismutase
110	37	43.0	86	1	D46335	W protein - Maedi/	183	36.5	42.4	203	2	T50044	superoxide dismutase
111	37	43.0	128	2	S63983	bile acid-binding	184	36.5	42.4	244	2	T50936	dehydrogenase/reductase
112	37	43.0	129	2	G81206	hypothetical protein	185	36.5	42.4	315	2	T29045	hypothetical protein
113	37	43.0	150	2	T39482	N-acetyltransferase	186	36.5	42.4	497	2	H83886	hypothetical protein
114	37	43.0	155	2	F83341	hypothetical protein	187	36.5	42.4	663	2	E83623	hypothetical protein
115	37	43.0	157	2	E83761	hypothetical protein	188	36.5	42.4	1254	1	JQ1978	structural polyprotein
116	37	43.0	181	2	A36367	ADP-ribosylation factor	189	36.5	42.4	1254	1	VHVVVE	structural polyprotein
117	37	43.0	181	2	A36367	ADP-ribosylation factor	190	36.5	42.4	1254	1	VHVVVE	structural polyprotein
118	37	43.0	191	2	S29008	ADP-ribosylation factor	191	36.5	42.4	1254	1	VHVVVT	structural polyprotein
119	37	43.0	224	2	B81783	hypothetical protein	192	36.5	42.4	1255	1	B44213	structural polyprotein
120	37	43.0	225	2	B81207	hypothetical protein	193	36.5	42.4	1255	1	D44213	structural polyprotein
121	37	43.0	277	2	S46330	hypothetical protein	194	36.5	42.4	2100	2	T03223	probable polyketide synthase
122	37	43.0	292	2	S71556	DNA-binding protein	195	36	41.9	94	2	B82518	hypothetical protein
123	37	43.0	297	2	F83433	DNA-3-methyladenine	196	36	41.9	174	2	D83692	hypothetical protein
124	37	43.0	361	2	F72862	hypothetical protein	197	36	41.9	184	2	F84904	hypothetical protein
125	37	43.0	362	2	T41842	AcMNPV orf101 - B0	198	36	41.9	204	2	S64838	hypothetical protein
126	37	43.0	369	2	G69254	Golichol-P-glucose	199	36	41.9	235	2	A81174	hypothetical protein
127	37	43.0	386	2	T12845	hypothetical protein	200	36	41.9	244	2	C75605	hypothetical protein
128	37	43.0	405	2	D83351	hypothetical protein	201	36	41.9	247	2	G64341	hypothetical protein
129	37	43.0	405	2	G75027	alanyl-tRNA synthetase	202	36	41.9	250	2	S07237	hypothetical protein
130	37	43.0	448	2	C95257	choline binding protein	203	36	41.9	255	2	A60944	ubiquinol-cytochrome c
131	37	43.0	448	2	C98122	choline binding protein	204	36	41.9	255	2	A60944	ubiquinol-cytochrome c
132	37	43.0	458	2	A46366	galactokinase (K)	205	36	41.9	282	2	Ar2329	hypothetical protein
133	37	43.0	484	2	F71061	hypothetical protein	206	36	41.9	288	2	D64694	hypothetical protein
134	37	43.0	494	2	AC3582	probable blue-copper	207	36	41.9	301	2	E95977	UTP-glucose-1-phosphate
135	37	43.0	517	1	S19243	tyrosinase-related	208	36	41.9	307	2	AE3054	UTP-glucose-1-phosphate
136	37	43.0	519	1	YRHUR2	dopachrome isomerase	209	36	41.9	307	2	H98231	exon protein (improvement)
137	37	43.0	522	2	I51245	tyrosinase related	210	36	41.9	311	2	S66011	transcription regulator
138	37	43.0	532	2	S51171	amino acid transporter	211	36	41.9	316	2	T19884	hypothetical protein
139	37	43.0	537	1	YRHUB6	tyrosinase-related	212	36	41.9	320	1	D49349	UTP-glucose-1-phosphate
140	37	43.0	537	1	YRMSB6	tyrosinase-related	213	36	41.9	320	2	T14624	hypothetical protein
141	37	43.0	546	1	C70393	probable adenylate	214	36	41.9	332	2	T45723	hypothetical protein
142	37	43.0	548	2	T15318	hypothetical protein	215	36	41.9	336	2	C71964	UDP-3-O-[3-hydroxy
143	37	43.0	552	2	E72283	alpha-galactosidase	216	36	41.9	336	2	D64544	UDP-3-O-[3-hydroxy
144	37	43.0	557	2	T27752	hypothetical protein	217	36	41.9	353	2	D69422	F420-nonreducing
145	37	43.0	593	2	JCS167	protein-tyrosine-phosphatase	218	36	41.9	356	1	S64902	probable sugar red
146	37	43.0	610	2	F95898	probable nodulatin	219	36	41.9	380	2	T18509	hypothetical protein
147	37	43.0	676	2	S41217	hypothetical protein	220	36	41.9	380	2	T21112	hypothetical protein
148	37	43.0	751	2	A81816	nitric oxide reductase	221	36	41.9	392	2	T24240	hypothetical protein
149	37	43.0	751	2	D81062	nitric oxide reductase	222	36	41.9	421	2	C96806	unknown protein T5
150	37	43.0	801	2	A47744	diacylglycerol kinase	223	36	41.9	423	2	T19145	hypothetical protein
151	37	43.0	818	2	T29560	hypothetical protein	224	36	41.9	431	2	E81357	glutamate-tRNA ligase
152	37	43.0	927	2	T38518	ribonuclease II RNase	225	36	41.9	441	2	JQ2191	nucleocapsid protein
153	37	43.0	946	2	F84280	ATP-dependent helicase	226	36	41.9	463	1	D70585	probable glycosylase
154	37	43.0	1024	2	G72041	exodeoxyribonuclease	227	36	41.9	480	2	C96744	hypothetical protein
155	37	43.0	1024	2	F86582	exodeoxyribonuclease	228	36	41.9	502	2	T49188	serine carboxypeptidase
156	37	43.0	1024	2	D81624	exodeoxyribonuclease	229	36	41.9	504	2	G87532	tryptophan halogenase
157	37	43.0	1032	2	F65071	hypothetical protein	230	36	41.9	512	2	A72866	major envelope glycoprotein
158	37	43.0	1032	2	C85943	probable oxidoreductase	231	36	41.9	529	1	VGNVAC	major envelope glycoprotein
159	37	43.0	1032	2	G91097	probable oxidoreductase	232	36	41.9	530	2	T41865	GP64/67 BFP orf128
160	37	43.0	1126	2	S04716	DNA-directed RNA polymerase	233	36	41.9	530	2	T40608	capA protein - Clostridium
161	37	43.0	1156	2	E69444	chromosome segregase	234	36	41.9	627	2	T49952	hypothetical protein
162	37	43.0	1171	2	T00380	KIAA0637 protein - human	235	36	41.9	635	2	S57714	capB protein - Clostridium
163	37	43.0	1208	2	AE1947	chromosome segregase	236	36	41.9	639	2	S23118	proprotein convertase
164	37	43.0	1386	2	T49316	profilaggrin related	237	36	41.9	671	2	S61099	leukotriene-A4 hydrolase
165	37	43.0	2491	1	A28372	insulin-like growth factor	238	36	41.9	1536	2	S59841	4-alpha-glucanotransferase
166	37	43.0	2499	1	A30788	mannose 6-phosphatase	239	36	41.9	2434	2	S44861	DNA topoisomerase
167	37	43.0	2629	2	T30987	telomerase-associated protein	240	36	41.9	2787	2	S45416	TEU1 protein - yeast
168	37	43.0	2824	2	T22759	hypothetical protein	241	35.5	41.3	50	2	B41308	vif protein - human
169	36.5	42.4	192	2	S43007	viral infectivity	242	35.5	41.3	185	2	T36546	hypothetical protein
170	36.5	42.4	192	2	S43006	viral infectivity	243	35.5	41.3	192	1	ASLJNA	vif protein - human
171	36.5	42.4	192	2	S42997	viral infectivity	244	35.5	41.3	192	1	ASLJNS	vif protein - human
172	36.5	42.4	192	2	S42958	viral infectivity	245	35.5	41.3	192	2	S42954	viral infectivity
173	36.5	42.4	192	2	S42968	viral infectivity	246	35.5	41.3	192	2	S42964	viral infectivity
174	36.5	42.4	192	2	S42941	viral infectivity	247	35.5	41.3	192	2	S42995	viral infectivity
175	36.5	42.4	192	2	S42992	viral infectivity	248	35.5	41.3	192	2	S42975	viral infectivity

249	35.5	41.3	192	2	S42946	viral infectivity	322	35	40.7	864	2	H85335	hypothetical prote
250	35.5	41.3	271	2	B83027	thiosulfate sulfur	323	35	40.7	864	2	T04518	hypothetical prote
251	35.5	41.3	328	1	Q0BEF6	HVLF4 protein - hu	324	35	40.7	900	2	G87431	pyruvate phosphate
252	35.5	41.3	647	2	B82579	acetyl coenzyme A	325	35	40.7	964	2	E71460	probable outer mem
253	35.5	41.3	873	2	JC4863	homeobox protein z	326	35	40.7	1056	2	E96748	hypothetical prote
254	35.5	41.3	1236	1	VHWWE	structural polypro	327	35	40.7	1182	2	T13952	membrane protein p
255	35	40.7	53	2	S39074	light-harvesting p	328	35	40.7	1245	2	E83110	exodeoxyribonuclea
256	35	40.7	107	2	A64643	hypothetical prote	329	35	40.7	1273	2	T00338	hypothetical prote
257	35	40.7	182	2	D83638	conserved hypotet	330	35	40.7	1369	2	T17504	hypothetical prote
258	35	40.7	196	2	T29343	hypothetical prote	331	35	40.7	1397	2	C64805	rhsc protein precu
259	35	40.7	205	2	E84464	hypothetical prote	332	35	40.7	1447	2	G86474	probable protein g
260	35	40.7	212	2	A81190	conserved hypotet	333	35	40.7	1630	2	S64403	ESPI protein - yea
261	35	40.7	212	2	G81913	hypothetical prote	334	35	40.7	1649	2	C96822	hypothetical prote
262	35	40.7	238	2	G90418	ABC transporter, A	335	35	40.7	2090	2	S26058	probable transform
263	35	40.7	249	2	E64404	hypothetical prote	336	35	40.7	3212	2	T24692	hypothetical prote
264	35	40.7	252	2	S77108	hypothetical prote	337	35	40.7	26926	1	T38344	tifin, cardiac mus
265	35	40.7	274	2	PRWVM	HIV-1 retropepsin	338	34.5	40.1	193	2	S25208	pruJ protein - Esc
266	35	40.7	279	2	JG0164	LIM protein, FHL4	339	34.5	40.1	203	2	A42710	superoxide dismuta
267	35	40.7	295	2	S46413	hypothetical prote	340	34.5	40.1	204	2	D71339	probable ribosomal
268	35	40.7	302	2	H69823	conserved hypotet	341	34.5	40.1	344	1	KHPGD	cathespain D (EC 3.
269	35	40.7	312	2	B96512	hypothetical prote	342	34.5	40.1	746	2	T24978	hypothetical prote
270	35	40.7	323	2	AC0117	hypothetical prote	343	34.5	40.1	748	2	S66129	disintegrin (EC 3.
271	35	40.7	330	2	H71309	probable asparagin	344	34.5	40.1	825	1	A40026	neurotrophin-3 rec
272	35	40.7	330	2	T49002	hypothetical prote	345	34.5	40.1	958	1	P18VCC	la protein - cowpe
273	35	40.7	332	2	S60935	hypothetical prote	346	34	39.5	74	2	T44088	probable transposa
274	35	40.7	333	2	A41881	collagenase PrtC (347	34	39.5	118	2	F34792	lg heavy chain pre
275	35	40.7	335	2	T20920	hypothetical prote	348	34	39.5	120	2	S36306	T-cell receptor de
276	35	40.7	340	2	T41729	probable adenosine	349	34	39.5	129	2	A75558	hypothetical prote
277	35	40.7	340	2	E69544	hypothetical prote	350	34	39.5	130	2	AH0463	probable acetyltra
278	35	40.7	351	2	T42538	adenosine kinase h	351	34	39.5	141	2	E87790	protein B0207.8 [i
279	35	40.7	351	2	H71432	probable glucosylt	352	34	39.5	146	2	C90477	quinol oxidase-2,
280	35	40.7	354	2	JQ0413	alkanal monooxygen	353	34	39.5	147	2	T01039	hypothetical prote
281	35	40.7	354	2	T33270	hypothetical prote	354	34	39.5	156	2	B85814	DNA mismatch endon
282	35	40.7	372	2	AE3191	conserved hypotet	355	34	39.5	156	2	B90966	patch repair prote
283	35	40.7	385	2	E86359	Similar to seed ma	356	34	39.5	156	2	AG0754	DNA mismatch endon
284	35	40.7	388	2	S04110	Integrase - Strept	357	34	39.5	156	2	JS0264	hypothetical prote
285	35	40.7	389	2	AH0547	methylicitrate synt	358	34	39.5	157	2	H81222	conserved hypotet
286	35	40.7	389	2	B90677	probable citrate s	359	34	39.5	158	2	A12899	hypothetical prote
287	35	40.7	399	2	E64760	citrate (si)-synth	360	34	39.5	158	2	B97675	hypothetical prote
288	35	40.7	399	2	E85527	hypothetical prote	361	34	39.5	178	2	A71730	ubiquinone biosynt
289	35	40.7	399	2	G83722	hypothetical prote	362	34	39.5	179	2	G97729	ubiquinone biosynt
290	35	40.7	393	2	S77691	probable finger pr	363	34	39.5	183	2	S31016	gene 71 protein -
291	35	40.7	438	2	B83017	probable MFS trans	364	34	39.5	184	2	S24450	terminase - phage
292	35	40.7	443	2	G90388	thermopsine precu	365	34	39.5	186	2	G97096	integrase/recombin
293	35	40.7	470	2	S33639	finger protein esc	366	34	39.5	191	2	G64017	hypothetical prote
294	35	40.7	471	2	T33997	hypothetical prote	367	34	39.5	193	2	E64488	hypothetical prote
295	35	40.7	472	2	S50859	P2x receptor - rat	368	34	39.5	196	2	D90389	conserved hypotet
296	35	40.7	477	2	T45722	hypothetical prote	369	34	39.5	198	2	S76099	probable orotate p
297	35	40.7	494	2	B95411	probable aldehyde	370	34	39.5	200	1	A54020	Crotalus neutraliz
298	35	40.7	496	2	S33791	ARS-binding protei	371	34	39.5	205	2	D84527	probable ADP-ribos
299	35	40.7	511	2	C95205	hypothetical prote	372	34	39.5	206	2	AC2443	probable phosphorib
300	35	40.7	512	1	ALBSL	alpha-amylase (EC	373	34	39.5	222	2	G86944	probable membrane
301	35	40.7	527	2	D84517	probable replicati	374	34	39.5	225	2	T34201	hypothetical prote
302	35	40.7	529	2	S46116	probable regulator	375	34	39.5	234	2	S26453	hypothetical prote
303	35	40.7	549	2	S49446	RING-finger protei	376	34	39.5	237	2	S26439	hypothetical prote
304	35	40.7	560	2	AB2437	NADH dehydrogenase	377	34	39.5	239	2	A11997	hypothetical prote
305	35	40.7	570	2	T30527	hypothetical prote	378	34	39.5	240	2	B82383	conserved hypotet
306	35	40.7	587	2	T32546	hypothetical prote	379	34	39.5	245	2	I57946	thiopurine methylt
307	35	40.7	594	2	A82913	hypothetical prote	380	34	39.5	259	2	T36003	hypothetical prote
308	35	40.7	617	2	E72803	gp31 protein - Myc	381	34	39.5	260	2	E83174	thioesterase limpo
309	35	40.7	626	2	T08001	hypothetical prote	382	34	39.5	261	2	C87204	outer membrane pro
310	35	40.7	655	2	B65217	hypothetical 73.7	383	34	39.5	270	2	A71907	transketolase homo
311	35	40.7	665	2	E86102	hypothetical prote	384	34	39.5	274	2	AH1203	hypothetical prote
312	35	40.7	665	2	B91262	hypothetical prote	385	34	39.5	277	2	AE1831	protein B0238.10 [
313	35	40.7	665	2	G82208	GGDEF family prote	386	34	39.5	278	2	F89044	hypothetical prote
314	35	40.7	713	2	F82506	probable TonB syst	387	34	39.5	287	2	D83856	hypothetical prote
315	35	40.7	716	2	D69855	conserved hypotet	388	34	39.5	292	2	F70877	hypothetical prote
316	35	40.7	770	2	A11769	autolysin, amidase	389	34	39.5	293	2	T06232	Ps16 protein - whe
317	35	40.7	780	2	I47038	vasopressin-activa	390	34	39.5	294	2	T05725	cp31AHV protein -
318	35	40.7	785	2	T19741	hypothetical prote	391	34	39.5	297	1	NBRF	apolipoprotein H p
319	35	40.7	802	2	C83588	probable hydroxama	392	34	39.5	303	2	S23780	nucleic acid-bindi
320	35	40.7	809	2	T32899	probable leukotrie	393	34	39.5	306	2	T09067	extensin-like prot
321	35	40.7	846	2	T27282	hypothetical prote	394	34	39.5	310	2	E64751	probable membrane

395	34	39.5	311	2	D82786	thiamin biosynthes	468	34	39.5	647	2	G75060	hydrogenase-4 comp
396	34	39.5	316	2	H82649	hypothetical prote	469	34	39.5	648	2	S59723	transcriptional acti
397	34	39.5	318	2	C70636	probable echA13 pr	470	34	39.5	661	2	T38176	hypothetical Myb f
398	34	39.5	320	2	AH1226	N-acetylmuramoyl-L	471	34	39.5	677	2	G69895	formate dehydrogen
399	34	39.5	328	2	G95187	conserved domain p	472	34	39.5	735	2	B87601	methyl-accepting c
400	34	39.5	333	2	A49405	protein kinase Pkn	473	34	39.5	754	2	E69745	hypothetical prote
401	34	39.5	334	2	A82751	ABC transporter At	474	34	39.5	829	2	I40014	sorbitol dehydroge
402	34	39.5	351	2	E84096	hypothetical prote	475	34	39.5	910	2	T29935	hypothetical prote
403	34	39.5	353	2	A86766	hypothetical prote	476	34	39.5	982	2	T19526	hypothetical prote
404	34	39.5	357	2	T23460	hypothetical prote	477	34	39.5	990	2	B49351	bacteriophage N4 a
405	34	39.5	374	2	F97309	uncharacterized co	478	34	39.5	990	2	H90703	bacteriophage N4 a
406	34	39.5	377	2	JC4612	actin - Chlamydomo	479	34	39.5	990	2	C85554	bacteriophage N4 a
407	34	39.5	377	2	SL4120	actin - Volvox car	480	34	39.5	995	2	C81593	polymorphic membra
408	34	39.5	379	2	T31154	hypothetical prote	481	34	39.5	998	2	S31735	NAD ADP-ribosyltra
409	34	39.5	381	2	AI0159	probable periplasm	482	34	39.5	1011	1	JH0581	NAD ADP-ribosyltra
410	34	39.5	383	2	T15698	hypothetical prote	483	34	39.5	1041	2	C83548	hypothetical prote
411	34	39.5	388	2	S36500	E2 protein - human	484	34	39.5	1054	2	A30239	hydroxymethylgluta
412	34	39.5	388	2	AI0208	oligogalacturonide	485	34	39.5	1068	1	A43322	1-phosphatidylinos
413	34	39.5	394	2	S15208	methane monooxygen	486	34	39.5	1068	1	T38110	1-phosphatidylinos
414	34	39.5	399	2	A33396	beta-N-acetylgluco	487	34	39.5	1107	2	S67381	tubulin-folding co
415	34	39.5	402	2	T13614	N-acetyltransferas	488	34	39.5	1204	2	B81947	probable exodeoxyr
416	34	39.5	403	2	D75333	conserved hypotet	489	34	39.5	1204	2	F81158	exodeoxyribonuclea
417	34	39.5	405	2	C75567	adenylosuccinate s	490	34	39.5	1206	2	E86445	hypothetical prote
418	34	39.5	407	2	E88968	protein T2B7.4 [i	491	34	39.5	1239	2	T13809	probable disintegr
419	34	39.5	408	2	AE0103	probable regulator	492	34	39.5	1327	2	D70759	probable otsB prot
420	34	39.5	412	2	AF2979	acyl-CoA dehydroge	493	34	39.5	1369	2	T43433	alpha-glucan synth
421	34	39.5	413	2	T43170	probable triacylgl	494	34	39.5	1385	2	T14158	neurexin IV - mous
422	34	39.5	417	2	F90916	probable transport	495	34	39.5	1474	1	MAHU	alpha-2-macroglobu
423	34	39.5	417	2	F64915	membrane protein y	496	34	39.5	1477	2	T18534	protein-tyrosine k
424	34	39.5	417	2	C85765	probable transport	497	34	39.5	1558	2	C89114	protein C37C3.6a [
425	34	39.5	422	2	A49518	kallistatin precur	498	34	39.5	1573	2	T21219	hypothetical prote
426	34	39.5	427	2	E83357	probable oxidoredu	499	34	39.5	1615	2	B49502	protein-tyrosine-p
427	34	39.5	432	2	S51474	hypothetical prote	500	34	39.5	1710	2	T14005	phospholipase D (E
428	34	39.5	436	2	AI1015	probable exported	501	34	39.5	1767	2	A49502	protein-tyrosine-p
429	34	39.5	443	2	T39540	triglyceride lipas	502	34	39.5	2167	2	T34395	hypothetical prote
430	34	39.5	448	2	F95348	nitric-oxide reduc	503	34	39.5	2406	2	A54148	odz protein - frui
431	34	39.5	448	2	AH3095	nitric oxide reduc	504	34	39.5	2469	2	H36812	hypothetical prote
432	34	39.5	448	2	B98191	nitric oxide reduc	505	34	39.5	2471	2	T42977	large tegument pro
433	34	39.5	449	2	AE3634	nitric-oxide reduc	506	34	39.5	2515	2	S47008	tenascin-like prot
434	34	39.5	451	2	JE0166	nitric-oxide reduc	507	33.5	39.0	180	2	E69269	molybdopterin oxid
435	34	39.5	451	2	G98303	hypothetical prote	508	33.5	39.0	192	2	T09442	vif protein - huma
436	34	39.5	463	2	D87012	probable glycol-tr	509	33.5	39.0	240	1	JQ0807	hydrogenase (EC 1.
437	34	39.5	463	2	T46165	pectate lyase-like	510	33.5	39.0	240	1	S53656	hydrogenase (EC 1.
438	34	39.5	471	2	T41318	probable glycoeyl	511	33.5	39.0	255	2	B70982	probable nei prote
439	34	39.5	476	2	F81340	glutamate-ammonia	512	33.5	39.0	322	1	W2MLE	E2 protein - human
440	34	39.5	481	2	B71929	glutamine syntheta	513	33.5	39.0	340	2	B88939	protein C05E4.11 [
441	34	39.5	481	2	H64583	glutamine syntheta	514	33.5	39.0	391	2	S39816	lysine acetyltrans
442	34	39.5	489	2	S62474	probable transcrip	515	33.5	39.0	567	2	JC5538	Rab geranylgeranyl
443	34	39.5	498	2	H85040	hypothetical prote	516	33.5	39.0	608	2	T25018	hypothetical prote
444	34	39.5	508	2	T36945	hypothetical prote	517	33.5	39.0	701	2	F90038	hypothetical prote
445	34	39.5	517	2	B71260	hypothetical prote	518	33.5	39.0	873	2	JC7079	homeobox protein Z
446	34	39.5	520	2	F70350	recombination prot	519	33.5	39.0	1094	2	E70697	probable arabinosy
447	34	39.5	533	1	YRMSCS	monophenol monooxy	520	33	38.4	63	2	AI2413	hypothetical prote
448	34	39.5	534	2	S60205	phosphonoacetaldeh	521	33	38.4	69	2	AC3565	hypothetical prote
449	34	39.5	544	2	S52081	diphosphate-fructo	522	33	38.4	75	2	I46471	alpha-actin - rabb
450	34	39.5	546	2	T23920	hypothetical prote	523	33	38.4	83	2	S77417	prochlorophyllid
451	34	39.5	547	2	A32803	glucan 1,4-alpha-m	524	33	38.4	87	2	S71587	ADP-ribosylation f
452	34	39.5	551	2	S05667	glucan 1,4-alpha-m	525	33	38.4	98	2	A39437	expopolysaccharide
453	34	39.5	555	2	D70102	pyrophosphate-fruc	526	33	38.4	98	2	F95975	posttranscription
454	34	39.5	564	2	AF2351	serine/threonine k	527	33	38.4	112	2	T44708	hypothetical prote
455	34	39.5	565	2	AI0479	probable membrane	528	33	38.4	112	2	T44903	hypothetical prote
456	34	39.5	566	2	S15387	malate synthase (E	529	33	38.4	120	2	E49590	Ig heavy chain V r
457	34	39.5	567	1	SYCNMU	malate synthase (E	530	33	38.4	124	2	C97067	hypothetical prote
458	34	39.5	568	1	SYKMA	malate synthase (E	531	33	38.4	129	2	B72332	transposase - ther
459	34	39.5	573	2	C71312	probable pyrophosp	532	33	38.4	129	2	D72205	transposase - ther
460	34	39.5	585	2	A46209	protein-tyrosine-p	533	33	38.4	140	2	S03109	actin - pin mould
461	34	39.5	593	1	JN0805	protein-tyrosine-p	534	33	38.4	149	2	T30925	hypothetical prote
462	34	39.5	595	1	A58651	protein-tyrosine-p	535	33	38.4	149	2	G91123	evolved beta-D-gal
463	34	39.5	597	1	A53593	protein-tyrosine-p	536	33	38.4	149	2	F85868	evolved beta-D-gal
464	34	39.5	599	2	A48863	limonene cyclase -	537	33	38.4	149	2	B65096	beta-galactosidase
465	34	39.5	622	2	G90250	glucan 1,4 alpha g	538	33	38.4	153	2	E87313	conserved hypotet
466	34	39.5	631	2	F81227	glucose inhibited	539	33	38.4	157	2	F71906	hypothetical prote
467	34	39.5	641	2	A42019	tyrosine-trNA liga	540	33	38.4	158	2	I49465	alpha-cardiac acti

541	33	38.4	158	2	P64607	hypothetical prote	614	33	38.4	274	2	AG1678	D-alanyl-D-alanine
542	33	38.4	160	2	S43605	R07E5.13 protein (615	33	38.4	278	2	C83280	probable transcrip
543	33	38.4	160	2	T16043	hypothetical prote	616	33	38.4	283	2	G82212	conserved hypothet
544	33	38.4	161	2	AG1073	conserved hypothet	617	33	38.4	283	2	E95416	restriction endonu
545	33	38.4	166	2	T37607	probable histone a	618	33	38.4	293	2	S50051	hypothetical prote
546	33	38.4	166	2	AB1150	transcription regu	619	33	38.4	296	2	T77768	hypothetical prote
547	33	38.4	166	2	AC1509	weakly transcripti	620	33	38.4	297	2	S76306	hypothetical prote
548	33	38.4	175	2	B53859	ADP-ribosylation f	621	33	38.4	300	2	G91288	hypothetical prote
549	33	38.4	175	2	B23741	ADP-ribosylation f	622	33	38.4	300	2	B86130	hypothetical prote
550	33	38.4	175	2	T31519	ADP-ribosylation f	623	33	38.4	300	2	S56545	fimbrial protein f
551	33	38.4	175	2	JC4950	ADP-ribosylation f	624	33	38.4	302	2	A32801	fimbrial adhesin p
552	33	38.4	175	4	I50632	hypothetical CPS1	625	33	38.4	305	2	T20685	hypothetical prote
553	33	38.4	176	2	AB0403	probable cytochrom	626	33	38.4	306	2	F83348	probable transcrip
554	33	38.4	180	1	S37599	ADP-ribosylation f	627	33	38.4	308	2	A03000	actin 3 - fruit fi
555	33	38.4	180	2	A53859	ADP-ribosylation f	628	33	38.4	308	2	AE2359	hypothetical prote
556	33	38.4	180	2	S57944	ADP-ribosylation f	629	33	38.4	312	2	S51440	hypothetical prote
557	33	38.4	180	2	A23741	ADP-ribosylation f	630	33	38.4	319	2	H69102	tyrosine-tRNA liga
558	33	38.4	180	2	T32978	ADP-ribosylation f	631	33	38.4	325	2	F94866	probable peroxidase
559	33	38.4	180	2	JC4949	ADP-ribosylation f	632	33	38.4	326	2	P50017	Ig gamma-1 chain C
560	33	38.4	180	2	JC4948	ADP-ribosylation f	633	33	38.4	326	2	T23139	hypothetical prote
561	33	38.4	180	2	I55371	ADP-ribosylation f	634	33	38.4	327	2	S11452	actin (clone 302)
562	33	38.4	181	2	A41570	ADP-ribosylation f	635	33	38.4	327	2	T30072	hypothetical prote
563	33	38.4	181	2	A33283	ADP-ribosylation f	636	33	38.4	328	2	S05430	actin beta - grass
564	33	38.4	181	2	A45422	ADP-ribosylation f	637	33	38.4	330	2	H64077	aspartate-ammonia
565	33	38.4	181	2	A36167	ADP-ribosylation f	638	33	38.4	331	2	D40649	D-2-hydroxy-acid d
566	33	38.4	181	2	T52339	ADP-ribosylation f	639	33	38.4	331	2	S24409	actin - brown alga
567	33	38.4	181	2	T52341	ADP-ribosylation f	640	33	38.4	332	2	S44206	hypothetical prote
568	33	38.4	181	2	S49325	ADP-ribosylation f	641	33	38.4	333	2	S15238	O-antigen acetatase
569	33	38.4	181	2	T15341	ADP-ribosylation f	642	33	38.4	333	2	T52594	squamosa promoter
570	33	38.4	181	2	S66337	ADP-ribosylation f	643	33	38.4	336	2	T04085	actin - maize (fra
571	33	38.4	181	2	JC4946	ADP-ribosylation f	644	33	38.4	339	2	T21473	hypothetical prote
572	33	38.4	181	2	JC4947	ADP-ribosylation f	645	33	38.4	347	2	A81988	probable N-acetyl-
573	33	38.4	181	2	JC4945	ADP-ribosylation f	646	33	38.4	347	2	D81043	N-acetyl-gamma-glu
574	33	38.4	181	2	T48021	ADP-ribosylation f	647	33	38.4	347	2	T37995	probable fatty aci
575	33	38.4	181	2	G96728	probable ADP-ribos	648	33	38.4	348	1	BVECMCB	mcrC protein - Esc
576	33	38.4	181	2	S28875	ADP-ribosylation f	649	33	38.4	349	2	B25819	actin, fetal skele
577	33	38.4	182	2	A49520	ADP-ribosylation f	650	33	38.4	350	2	A54420	beta-galactoside a
578	33	38.4	182	2	C49993	ADP-ribosylation f	651	33	38.4	357	1	G69290	probable hexosyltr
579	33	38.4	183	2	D49993	ADP-ribosylation f	652	33	38.4	361	2	S68089	actin 2 - Arabidop
580	33	38.4	186	2	E69537	conserved hypothet	653	33	38.4	362	2	A26559	actin type 5, cyto
581	33	38.4	188	2	E86368	FS08.5 protein - A	654	33	38.4	362	2	S68090	actin 8 - Arabidop
582	33	38.4	188	2	T48640	ADP-ribosylation f	655	33	38.4	365	2	A37431	actin, type 1 - Em
583	33	38.4	190	2	H83172	hypothetical prote	656	33	38.4	365	2	S49007	actin - Pythium ir
584	33	38.4	190	2	A82343	conserved hypothet	657	33	38.4	367	2	UT0596	actin Ardd - slime
585	33	38.4	192	1	C40001	vif protein - huma	658	33	38.4	370	2	A29664	actin beta - bovin
586	33	38.4	192	2	S42988	viral infectivity	659	33	38.4	374	1	ATB08	actin gamma - bovi
587	33	38.4	192	2	T39367	hypothetical prote	660	33	38.4	374	1	ATBOG	gamma-actin - huma
588	33	38.4	193	2	AG3371	hypothetical prote	661	33	38.4	374	2	JC5818	pectate lyase (EC
589	33	38.4	194	2	A70838	hypothetical prote	662	33	38.4	375	1	W2WC6C	actin, aortic smoo
590	33	38.4	195	2	S39777	actin beta - pig (663	33	38.4	375	1	ATBOSM	actin, skeletal mu
591	33	38.4	195	2	S20097	actin 85c - potato	664	33	38.4	375	1	ATRB	actin beta - rat
592	33	38.4	197	2	S36453	ADP-ribosylation f	665	33	38.4	375	1	ATRTC	actin beta - cytosk
593	33	38.4	201	2	A11106	conserved hypothet	666	33	38.4	375	1	A48324	actin - Acanthamo
594	33	38.4	201	2	AE1468	hypothetical prote	667	33	38.4	375	1	ATAX	actin - yeast (Sac
595	33	38.4	204	2	T04658	hypothetical prote	668	33	38.4	375	1	ATBY	actin beta - chick
596	33	38.4	205	2	T07423	actin - Chlorella	669	33	38.4	375	1	ATCHB	actin beta - slime mold
597	33	38.4	213	2	A61043	actin CA15 - sea s	670	33	38.4	375	1	ATDO	actin beta - human
598	33	38.4	213	2	D69938	hemolysin III homo	671	33	38.4	375	1	ATHUB	actin gamma 1 - hu
599	33	38.4	225	2	H75571	conserved hypothet	672	33	38.4	375	1	ATHUG	actin gamma 1 - hu
600	33	38.4	225	2	S73585	MG068 homolog D02	673	33	38.4	375	1	ATMSB	actin beta - mouse
601	33	38.4	231	2	H86463	Fl2G12.17 protein	674	33	38.4	375	1	ATMSG	actin gamma - mous
602	33	38.4	232	2	A42095	floral homeotic pr	675	33	38.4	375	1	ATRB	actin beta, non-mu
603	33	38.4	233	2	S63660	hypothetical prote	676	33	38.4	375	1	ATZM1	actin - maize
604	33	38.4	241	2	D71167	hypothetical prote	677	33	38.4	375	1	JS0702	actin - yeast (Sac
605	33	38.4	247	2	G82202	hypothetical prote	678	33	38.4	375	1	S11222	actin gamma, cytos
606	33	38.4	252	2	AF1861	ATP-binding protei	679	33	38.4	375	2	AJ2798	actin - yeast (Klu
607	33	38.4	258	2	E90126	hypothetical prote	680	33	38.4	375	2	A26836	actin beta - fission ye
608	33	38.4	260	2	I51544	MHC class II beta-	681	33	38.4	375	2	A55001	actin beta - goose
609	33	38.4	261	2	E72735	hypothetical prote	682	33	38.4	375	2	A54728	actin alpha, cardi
610	33	38.4	273	2	B84868	probable endociti	683	33	38.4	375	2	S33386	actin, cytosolic (
611	33	38.4	273	2	B64608	outer membrane pro	684	33	38.4	375	2	UT0385	actin gamma - Emer
612	33	38.4	274	2	C71300	hypothetical prote	685	33	38.4	375	2	S78997	actin 1 - Pneumocy
613	33	38.4	274	2	AG1306	D-alanyl-D-alanine	686	33	38.4	375	2	T25272	hypothetical prote

687	33	38.4	375	2	S71125	actin beta-2, cyto	760	33	38.4	377	1	ATHU	actin alpha 1, ske
688	33	38.4	375	2	S71126	actin beta-1, cyto	761	33	38.4	377	1	ATHUC	actin, cardiac mus
689	33	38.4	375	2	S71126	actin beta, cyto	762	33	38.4	377	1	ATHUSM	actin alpha 2, aor
690	33	38.4	375	2	S42103	actin - Puccinia g	763	33	38.4	377	1	ATMUM1	actin - Arabidopsi
691	33	38.4	375	2	S03126	actin - imperfect	764	33	38.4	377	1	ATRBMS	actin alpha, smoot
692	33	38.4	375	2	S70377	actin - Phaffia rh	765	33	38.4	377	1	ATRT	actin, skeletal mu
693	33	38.4	375	2	E82681	conserved hypothet	766	33	38.4	377	1	ATRZ1	actin 1 - rice
694	33	38.4	376	1	ATFF87	actin 7 - fruit fl	767	33	38.4	377	1	B29686	actin alpha, card
695	33	38.4	376	1	ATFSY3	actin - soybean	768	33	38.4	377	2	B24848	actin alpha-3, ske
696	33	38.4	376	1	A43552	actin gamma, cytos	769	33	38.4	377	2	A24848	actin alpha-1, car
697	33	38.4	376	1	ATAXE	actin - Entamoeba	770	33	38.4	377	2	A29686	actin alpha-2, ske
698	33	38.4	376	1	ATCHSM	actin gamma, smoot	771	33	38.4	377	2	JCS301	skeletal alpha-act
699	33	38.4	376	1	ATFF8	actin 8 - fruit fl	772	33	38.4	377	2	S31933	actin - common tob
700	33	38.4	376	1	ATFY	actin - slime mold	773	33	38.4	377	2	S20093	actin 101 - potato
701	33	38.4	376	1	ATRFZ	actin 3 - rice	774	33	38.4	377	2	S20098	actin 97 - potato
702	33	38.4	376	1	ATRZ7	actin 7 - rice	775	33	38.4	377	2	S20096	actin 75 - potato
703	33	38.4	376	1	ATUBS	actin Cyl - sea ur	776	33	38.4	377	2	S20094	actin 58 - potato
704	33	38.4	376	1	E69957	gamma-D-glutamyl-L	777	33	38.4	377	2	S20095	actin 71 - potato
705	33	38.4	376	2	B23412	actin 12 - slime m	778	33	38.4	377	2	JE0147	actin 1 - sorghum
706	33	38.4	376	2	A48449	Actin-1A - nematod	779	33	38.4	377	2	S71120	actin alpha, card
707	33	38.4	376	2	A27724	actin 1 - Trypanos	780	33	38.4	377	2	S71118	actin alpha-1, ske
708	33	38.4	376	2	B27724	actin 2 - Trypanos	781	33	38.4	377	2	S71119	actin alpha-2, ske
709	33	38.4	376	2	A45634	actin - Cryptospor	782	33	38.4	377	2	S68107	actin 7 - Arabidop
710	33	38.4	376	2	A54496	actin I - malaria	783	33	38.4	377	2	S68112	actin 3 [imported]
711	33	38.4	376	2	A54509	actin II - malaria	784	33	38.4	377	2	S68110	actin 12 - Arabido
712	33	38.4	376	2	A29407	actin - Tetrahymen	785	33	38.4	377	2	S68108	actin 4 - Arabido
713	33	38.4	376	2	A40261	actin gamma, enter	786	33	38.4	377	2	S68109	actin 11 - Arabido
714	33	38.4	376	2	C23412	actin 3-sub1 - sli	787	33	38.4	377	2	S58316	actin - garden pea
715	33	38.4	376	2	A25084	actin 15 - slime m	788	33	38.4	377	2	T51177	actin [imported] -
716	33	38.4	376	2	A44940	actin - pork tapew	789	33	38.4	377	2	T51178	actin AC12 [import
717	33	38.4	376	2	A32788	actin gamma, smoot	790	33	38.4	377	2	T51181	actin 1 [imported]
718	33	38.4	376	2	A31375	actin, smooth musc	791	33	38.4	377	2	T51183	actin isoform B [i
719	33	38.4	376	2	A25135	actin A3, cytosoli	792	33	38.4	377	2	T51180	actin [imported] -
720	33	38.4	376	2	S04538	actin 87E - fruit	793	33	38.4	377	2	T51175	actin [imported] -
721	33	38.4	376	2	JC1246	actin - fruit fly	794	33	38.4	377	2	T51184	actin [imported] -
722	33	38.4	376	2	S12628	actin - malaria pa	795	33	38.4	377	2	T51179	actin [imported] -
723	33	38.4	376	2	S07284	actin - Tetrahymen	796	33	38.4	377	2	T51176	actin [imported] -
724	33	38.4	376	2	JS0189	actin, cytosolic -	797	33	38.4	377	2	T51182	actin [imported] -
725	33	38.4	376	2	JS0190	actin, muscle - st	798	33	38.4	378	1	ATSY1	actin 1 - soybean
726	33	38.4	376	2	S07288	actin 15A - sea ur	799	33	38.4	378	2	H84849	probable actin [im
727	33	38.4	376	2	S09578	actin - sea urchin	800	33	38.4	379	1	ATRZ2	actin 2 - rice
728	33	38.4	376	2	JQ0154	actin - Hydra atte	801	33	38.4	379	2	S33387	actin, muscle - se
729	33	38.4	376	2	JN0832	actin (clone gen3)	802	33	38.4	380	2	D23412	actin 3-sub2 - sli
730	33	38.4	376	2	JN0833	actin (clones Ia a	803	33	38.4	380	2	S07002	actin 1 - carrot
731	33	38.4	376	2	JN0833	actin - Achlya bis	804	33	38.4	380	2	A83458	class I histocampa
732	33	38.4	376	2	S24408	actin A - Phytopt	805	33	38.4	381	2	S35940	bacteroid developm
733	33	38.4	376	2	JE0414	actin 4 - Caenorha	806	33	38.4	382	2	AG2859	hypothetical prote
734	33	38.4	376	2	S27135	actin 1 and actin	807	33	38.4	383	2	A12748	probable ATP-bind
735	33	38.4	376	2	S16710	actin 2 - Caenorha	808	33	38.4	383	2	H97529	hypothetical prote
736	33	38.4	376	2	T24448	hypothetical prote	809	33	38.4	386	2	T06788	actin - garden pea
737	33	38.4	376	2	JCS228	actin 2 - earthwor	810	33	38.4	387	2	D88968	protein T27B7.3 li
738	33	38.4	376	2	JCS227	actin 1 - earthwor	811	33	38.4	389	2	T32516	hypothetical prote
739	33	38.4	376	2	S12730	actin - California	812	33	38.4	394	2	A81789	probable monooxyge
740	33	38.4	376	2	S43509	actin - California	813	33	38.4	394	2	F81212	UbliH family protei
741	33	38.4	376	2	S11450	actin (clone 205)	814	33	38.4	395	2	A13402	salicylate 1-monoo
742	33	38.4	376	2	S11451	actin (clone 211)	815	33	38.4	400	2	S07733	NADH2 dehydrogenas
743	33	38.4	376	2	S11453	actin (clone 403)	816	33	38.4	402	1	RERTK	renin (EC 3.4.23.1
744	33	38.4	376	2	S49481	actin 5 - Atlantic	817	33	38.4	403	2	A72523	probable acyl-CoA
745	33	38.4	376	2	S49479	actin 11 - Atlanti	818	33	38.4	413	2	H86823	ammonium transport
746	33	38.4	376	2	S49480	actin 3 - Atlantic	819	33	38.4	419	2	E83904	hypothetical prote
747	33	38.4	376	2	S71123	actin alpha-anomal	820	33	38.4	422	2	E97636	bacteroid develop
748	33	38.4	376	2	S01077	actin beta, cytosk	821	33	38.4	423	2	A64486	dihydroorotase (EC
749	33	38.4	376	2	S25488	actin 1 - garden p	822	33	38.4	427	2	A64329	phosphoryruvate hy
750	33	38.4	376	2	S26435	actin 2 - garden p	823	33	38.4	427	2	T45915	actin (ACT3) - Ara
751	33	38.4	376	2	S07639	actin - Yeast (Can	824	33	38.4	429	2	T14237	hypothetical prote
752	33	38.4	376	2	JCS750	actin A4 - silkwor	825	33	38.4	431	2	T41560	phenylalanyl-trna
753	33	38.4	376	2	S09059	actin A1 - silkwor	826	33	38.4	434	2	S77330	hypothetical prote
754	33	38.4	376	2	S07382	actin A2 - silkwor	827	33	38.4	434	2	G70011	conserved hypothet
755	33	38.4	377	1	A25719	actin alpha, aorti	828	33	38.4	435	2	G87334	acyl-CoA dehydroge
756	33	38.4	377	1	A22224	actin alpha, vascu	829	33	38.4	436	2	G85749	aminoacylase (EC 3
757	33	38.4	377	1	A23022	actin, cardiac mus	830	33	38.4	441	2	T35169	probable glucosida
758	33	38.4	377	1	A24904	actin alpha, skele	831	33	38.4	441	2	E64883	probable aminohydr
759	33	38.4	377	1	ATCH	actin alpha, skele	832	33	38.4	441	2	B90869	probable aminohydr

833	33	38.4	443	2	T21598	hypothetical prote	906	33	38.4	761	2	A10368	malate dehydrogena
834	33	38.4	447	2	C85085	hypothetical prote	907	33	38.4	762	2	JC7174	N,N-dimethylformam
835	33	38.4	447	2	AD2474	hypothetical prote	908	33	38.4	765	2	JC7498	hypothetical prote
836	33	38.4	454	2	T25203	probable membrane-	909	33	38.4	794	2	T39171	probable peroxisom
837	33	38.4	462	2	G01804	interleukin 3-regu	910	33	38.4	800	2	T26683	hypothetical prote
838	33	38.4	463	2	T09243	dnak-type molecula	911	33	38.4	816	2	H85028	probable GTP pyrop
839	33	38.4	463	2	D87624	sodium-galactoside	912	33	38.4	828	1	D39142	outer membrane ush
840	33	38.4	473	1	WMBE51	Uti10 protein - hum	913	33	38.4	837	1	S54429	outer membrane ush
841	33	38.4	484	2	AD0431	xylokinase (EC 2	914	33	38.4	855	2	JC7731	membrane-bound arg
842	33	38.4	493	1	ACMSE	nicotinic acetylch	915	33	38.4	859	1	S65938	nitrate reductase
843	33	38.4	493	2	T34453	hypothetical prote	916	33	38.4	862	2	S64821	probable protein k
844	33	38.4	495	2	S71900	RNA-directed DNA p	917	33	38.4	871	2	G84601	protein kinase hom
845	33	38.4	501	2	S16711	ABC1 protein precu	918	33	38.4	878	2	T08559	TonB-dependent rec
846	33	38.4	505	2	T22558	hypothetical prote	919	33	38.4	937	2	G87640	nucleoporin Nup98
847	33	38.4	506	2	T07942	probable squalene	920	33	38.4	937	2	A56517	traG protein - Esc
848	33	38.4	511	2	T15629	hypothetical prote	921	33	38.4	938	2	S20480	probable nuclear m
849	33	38.4	514	1	ALBSN	alpha-amylase (EC	922	33	38.4	955	2	T39765	two-component hybr
850	33	38.4	515	2	JC5458	inulinase (EC 3.2.	923	33	38.4	1009	2	C64483	hypothetical prote
851	33	38.4	516	2	D64410	replication factor	924	33	38.4	1060	1	P2XRA4	outer capsid prote
852	33	38.4	516	2	H72427	DNA mismatch repai	925	33	38.4	1090	2	D72048	pbp2-transglycolas
853	33	38.4	516	2	JE0301	inulinase (EC 3.2.	926	33	38.4	1113	2	D86142	hypothetical prote
854	33	38.4	529	1	YRHU1	monophenol monooxy	927	33	38.4	1186	2	AG1928	two-component hybr
855	33	38.4	529	2	B49993	glycylpeptide N-te	928	33	38.4	1217	2	F97177	alpha-glucosidase
856	33	38.4	530	2	AC2085	phosphodiesterase/	929	33	38.4	1232	2	T43027	neural cell adhesi
857	33	38.4	532	2	JC1392	monophenol monooxy	930	33	38.4	1271	2	D64237	hypothetical prote
858	33	38.4	533	2	T32389	hypothetical prote	931	33	38.4	1289	1	GUBPT4	proximal tail fibe
859	33	38.4	535	2	G95155	hypothetical prote	932	33	38.4	1313	2	T29027	hypothetical prote
860	33	38.4	535	2	B98022	hypothetical prote	933	33	38.4	1397	2	T51292	Dna2p - fission ye
861	33	38.4	537	2	H88087	protein B0454.4 [i	934	33	38.4	1398	2	T39568	hypothetical helic
862	33	38.4	539	2	S53529	monophenol monooxy	935	33	38.4	1449	2	T30857	glucosyltransferas
863	33	38.4	545	2	T02279	hypothetical prote	936	33	38.4	1449	2	T30552	glucosyltransferas
864	33	38.4	555	1	S17502	inulinase (EC 3.2.	937	33	38.4	1518	2	A44811	glucosyltransferas
865	33	38.4	556	1	S31330	inulinase (EC 3.2.	938	33	38.4	1562	2	T17411	polyketide synthas
866	33	38.4	562	2	S16594	regulatory protein	939	33	38.4	1790	1	S27772	vitellogenin precu
867	33	38.4	567	2	D84400	phenylalanyl-tRNA	940	33	38.4	1928	2	JS0610	beta-galactosidase
868	33	38.4	584	2	H86631	C7085 hypothetical	941	33	38.4	2470	2	I50726	cation-independent
869	33	38.4	584	2	I50431	s-glycerin precurs	942	33	38.4	2817	2	B37033	uncharacterized pr
870	33	38.4	584	2	A72092	conserved hypothet	943	33	38.4	2958	2	S64921	probable membrane
871	33	38.4	586	2	S65802	transcription acti	944	33	38.4	3034	2	T14119	seven-pass transme
872	33	38.4	586	2	T15259	hypothetical prote	945	33	38.4	3305	2	T18358	apolipoprotein prec
873	33	38.4	598	2	S51456	probable membrane	946	32.5	37.8	133	2	B84087	hypothetical prote
874	33	38.4	598	2	S66954	probable membrane	947	32.5	37.8	183	1	I64006	hypothetical prote
875	33	38.4	599	2	JN0818	transferrin-bindin	948	32.5	37.8	192	2	S42944	viral infectivity
876	33	38.4	599	2	S67084	probable membrane	949	32.5	37.8	192	2	S43003	viral infectivity
877	33	38.4	605	2	G85651	probable membrane	950	32.5	37.8	192	2	S42950	conserved hypothet
878	33	38.4	605	2	D90791	probable membrane	951	32.5	37.8	192	2	AB0726	conserved hypothet
879	33	38.4	610	2	A41388	Lc regulatory prot	952	32.5	37.8	199	2	AC3120	superoxide dismuta
880	33	38.4	612	2	S06072	regulatory protein	953	32.5	37.8	200	1	DSHSNH	superoxide dismuta
881	33	38.4	615	2	T01422	GTP pyrophosphokin	954	32.5	37.8	200	2	B84288	superoxide dismuta
882	33	38.4	616	2	S20089	gene SN protein -	955	32.5	37.8	202	2	T50047	superoxide dismuta
883	33	38.4	620	2	S52494	protein kinase hom	956	32.5	37.8	202	2	A83826	superoxide dismuta
884	33	38.4	627	1	S27393	sphingomyelin phos	957	32.5	37.8	211	2	B75415	superoxide dismuta
885	33	38.4	632	2	AE3560	proline/betaine tr	958	32.5	37.8	226	2	B69709	superoxide dismuta
886	33	38.4	634	2	T01408	inclusion protein	959	32.5	37.8	240	2	D98167	superoxide dismuta
887	33	38.4	649	2	G71530	probable ATP synth	960	32.5	37.8	261	2	T11512	cytochrome-c oxida
888	33	38.4	649	2	C81687	ATP synthase, chai	961	32.5	37.8	296	2	JC7283	hydroxyarylamine s
889	33	38.4	651	2	D72605	probable long-chai	962	32.5	37.8	296	2	UC7282	hydroxyarylamine s
890	33	38.4	659	2	AB0731	probable hydrolase	963	32.5	37.8	296	2	D49539	xyloglucan endo-1,
891	33	38.4	660	2	E81549	ATP synthase, chai	964	32.5	37.8	339	1	NCBPX4	exonuclease 47 (EC
892	33	38.4	660	2	C86502	ATP synthase subun	965	32.5	37.8	365	2	AF3217	DNA-damage-inducib
893	33	38.4	660	2	E72121	ATP synthase chain	966	32.5	37.8	412	2	AG0177	probable hydroxyme
894	33	38.4	662	2	S50319	hypothetical F-box	967	32.5	37.8	428	2	B82248	long-chain fatty a
895	33	38.4	663	1	TMVVR	protein-tyrosine k	968	32.5	37.8	449	2	P83627	hypothetical prote
896	33	38.4	671	2	S51599	Om(2D) protein - f	969	32.5	37.8	555	2	AE2147	cytochrome c oxida
897	33	38.4	694	2	S71786	wingless receptor	970	32.5	37.8	648	2	SI0869	enterotoxin A - Cl
898	33	38.4	704	2	S46000	probable membrane	971	32.5	37.8	682	2	JQ0420	beta-1,3-glucanase
899	33	38.4	707	2	T09340	hypothetical prote	972	32.5	37.8	787	1	JDVLW2	DNA-directed DNA p
900	33	38.4	721	2	H82528	L-ascorbate oxidas	973	32.5	37.8	859	2	AC2089	adenylate cyclase
901	33	38.4	727	2	S47857	basic protein, cyt	974	32.5	37.8	1624	2	C70867	probable Helix-tur
902	33	38.4	746	2	S71892	RNA-directed DNA p	975	32	37.2	19	2	S43657	hsp90 protein homo
903	33	38.4	751	1	S26071	photosystem I prot	976	32	37.2	68	2	AG2291	hypothetical prote
904	33	38.4	757	2	JS0198	genome polyprotein	977	32	37.2	70	2	T04408	actin - barley (fr
905	33	38.4	760	2	A99233	hypothetical prote	978	32	37.2	81	2	A24522	mitochondrial prot

979 32 37.2 85 2 B47354 ysdF protein - Sal
980 32 37.2 86 2 G82801 hypothetical prote
981 32 37.2 89 2 D64465 hypothetical prote
982 32 37.2 91 2 I48658 dm1 antigen - mous
983 32 37.2 94 2 T29829 hypothetical prote
984 32 37.2 101 2 B38189 protein-tyrosine-p
985 32 37.2 107 2 AC3315 hypothetical membr
986 32 37.2 108 2 B30352 protein-lysine 6-o
987 32 37.2 109 2 PC2185 heat shock protein
988 32 37.2 111 2 F69779 transcription regu
989 32 37.2 117 2 F83449 conserved hypotet
990 32 37.2 119 2 S53764 hypothetical prote
991 32 37.2 120 2 AH0858 probable 6-pyruvoy
992 32 37.2 122 2 G82217 probable 6-pyruvoy
993 32 37.2 125 2 B83857 chorismate mutase
994 32 37.2 129 2 S76606 hypothetical prote
995 32 37.2 130 2 AI2467 50S ribosomal prot
996 32 37.2 135 2 T40220 hypothetical prote
997 32 37.2 136 2 B86628 prophage psi prote
998 32 37.2 136 2 AI0124 probable prepilin
999 32 37.2 140 2 S54084 probable membrane
1000 32 37.2 140 2 C95999 hypothetical prote

ALIGNMENTS

RESULT 1
F71190
A;Title: Probable chromosome assembly protein - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C;Accession: F71190
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Hakiwa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
C;Accession: F71190
A;Status: Preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1179 <KAW>
A;Cross-references: UNIPROT:O59462; UNIPARC:UPI000006688D; GB:AF000007; NID:g3236134; PT
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1798

Query Match 59.3%; Score 51; DB 2; Length 1179;
Best Local Similarity 81.8%; Pred. No. 2.5;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 HYLLNGRTATR 14
|||:|||||
Db 113 HYWLNGRRATR 123

RESULT 2
JC5251
A;Title: beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Oct-2004
C;Accession: JC5251; G01021
R;Kim, Y.J.; Kim, K.S.; Kim, S.H.; Kim, C.H.; Ko, J.H.; Choe, I.S.; Tsuji, S.; Lee, Y.C.
Biochem. Biophys. Res. Commun. 228, 324-327, 1996
A;Title: Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialyltra
A;Reference number: JC5251; MUID:97079181; PMID:8920913
C;Accession: JC5251
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-350 <KIM>
A;Cross-references: UNIPROT:Q16842; UNIPARC:UPI0000001C5C; GB:U63090; NID:g1773282; PIDN
A;Experimental source: liver

R;Giordanengo, V.
submitted to the EMBL Data Library, March 1996
A;Reference number: H00561
A;Accession: G01021
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-350 <GIO>
A;Cross-references: UNIPARC:UPI0000001C5C; EMBL:X96667; NID:g1235530
C;Comment: This enzyme catalyzes the transfer of sialic acid from CMP-NeuAc to the termin
substrate preference for glycolipid than for O-linked oligosaccharides of glycoproteins.
C;Genetics:
A;Gene: ST3(0)-II
C;Superfamily: sialyltransferase
C;Keywords: glycosyltransferase
Query Match 53.5%; Score 46; DB 2; Length 350;
Best Local Similarity 63.6%; Pred. No. 4.7;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
||||:|||||
Db 307 WHYYWNNRYA 317

RESULT 3
B54420
A;Title: beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) ST3GALNA.2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 05-Oct-2004
C;Accession: B54420
R;Lee, Y.C.; Kojima, N.; Wada, E.; Kurosawa, N.; Nakaoka, T.; Hamamoto, T.; Tsuji, S.
J. Biol. Chem. 269, 10028-10033, 1994
A;Title: Cloning and expression of cDNA for a new type of Galbeta1, eGalNAc alpha2,3-sialy
A;Reference number: A54420; MUID:94193584; PMID:8144500
C;Accession: B54420
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-350 <LEE>
A;Cross-references: UNIPROT:Q11205; UNIPARC:UPI000013596B; GB:X76986; NID:g475225; PIDN:
C;Superfamily: sialyltransferase
C;Keywords: glycosyltransferase; transmembrane protein

Query Match 53.5%; Score 46; DB 2; Length 350;
Best Local Similarity 63.6%; Pred. No. 4.7;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
||||:|||||
Db 307 WHYYWNNRYA 317

RESULT 4
T01245
A;Title: N-acetyltransferase homolog Fl6M14.6 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C;Accession: T01245; C84801
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC Fl6M14 genomic sequence.
A;Reference number: Z14213
A;Accession: T01245
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-190 <ROU>
A;Cross-references: UNIPROT:O80438; UNIPARC:UPI0000009EFCB; EMBL:AC003028; NID:g3335356; I
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84801
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-190 <STO>
A;Cross-references: UNIPARC:UPI000009EFCB; GB:AR002093; NID:G3335361; PIDN:AAC27162.1; C
C;Genetics:
A;Gene: At2g38130; F16M14.6
A;Map position: 2
A;Introns: 60/3; 101/1; 119/3

Query Match 52.3%; Score 45; DB 2; Length 190;
Best Local Similarity 66.7%; Pred. No. 3.5;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHYYLNGRTATR 14
:|||||
Db 147 YHYLNGMDAPR 158

RESULT 5
AC2030
hypothetical protein alr1793 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 31-Dec-2004
C;Accession: AC2030
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-290 <KUR>
A;Cross-references: UNIPROT:Q8YM23; UNIPARC:UPI000000CE1F1; GB:BA000019; PIDN:BAB73492.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr1793
C;Superfamily: nodulation protein nodB

Query Match 51.2%; Score 44; DB 2; Length 290;
Best Local Similarity 53.8%; Pred. No. 8.2;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTAT 13
|||||
Db 159 TWHYRMRMEAT 171

RESULT 6
S55675
Gal-beta-1,3GalNac alpha-2,3-sialyltransferase - chicken
C;Species: Gallus gallus (chicken)
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 05-Oct-2004
C;Accession: S55675
R;Kurosawa, N.; Hamamoto, T.; Inoue, M.; Tsuji, S.
Biochim. Biophys. Acta 1244, 216-222, 1995
A;Title: Molecular cloning and expression of chick Gal-beta-1,3GalNac alpha-2,3-sialyltr
A;Reference number: S55675; MUID:95284088; PMID:7766661
A;Accession: S55675
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-342 <KUR>
A;Cross-references: UNIPROT:Q11200; UNIPARC:UPI00000135969; GB:X80503; NID:g975654; PIDN:
C;Superfamily: sialyltransferase

Query Match 50.0%; Score 43; DB 2; Length 342;
Best Local Similarity 54.5%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
|||||

Db 299 WHYYWENASA 309

RESULT 7
D85330
hypothetical protein AT4g28370 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: D85330
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: D85330
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-479 <STO>
A;Cross-references: UNIPROT:O49446; UNIPARC:UPI000009FD49; GB:NC_001368; NID:g7369691; P1
C;Genetics:
A;Gene: AT4g28370
A;Map position: 4

Query Match 50.0%; Score 43; DB 2; Length 479;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHYYLNGRTATR 14
|||||
Db 330 HPYYILGWTATR 341

RESULT 8
JC7506
heparanase protein 2a - human
C;Species: Homo sapiens (man)
C;Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C;Accession: JC7506
R;McKenzie, E.; Iysson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Pat
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A;Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me
A;Reference number: JC7506
A;Accession: JC7506
A;Molecule type: mRNA
A;Residues: 1-480 <MCK>
A;Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI0000003E88A; GB:AF282885
C;Comment: This protein, an intracellular membrane-bound enzyme, has biological and therai
therapies.
C;Genetics:
A;Gene: hpa2a
A;Map position: 10q23-10q24
C;Keywords: heparin binding; membrane bound

Query Match 50.0%; Score 43; DB 2; Length 480;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TWHYYLNGR 10
|||||
Db 220 TWQHCYIDGR 229

RESULT 9
T04607
hypothetical protein F2009.50 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-May-1999
C;Accession: T04607
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.
submitted to the Protein Sequence Database, October 1998
A;Reference number: Z15380
A;Accession: T04607
A;Molecule type: DNA
A;Residues: 1-481 <BEV>

A;Cross-references: UNIPARC:UPI0000179E9C; EMBL:AL021749
 A;Experimental source: cultivar Columbia; BAC clone F2009
 C;Genetics:
 A;Map position: 4
 A;Introns: 4/3; 29/1; 82/3; 106/1; 128/2; 177/1; 226/3; 323/3; 420/1; 457/1
 A;Note: F2009.50

Query Match 50.0%; Score 43; DB 2; Length 481;
 Best Local Similarity 66.7%; Pred. No. 21;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 3 HHYYLNGRTATR 14
 | ||| | ||||
 DB 259 HPHYILGWTATR 270

RESULT 10
 B75150
 chromosome segregation protein (smc1) PAB2109 - Pyrococcus abyssi (strain Orsay)
 C;Species: Pyrococcus abyssi
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
 C;Accession: B75150
 R;Anonymous; Genoscope
 A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
 A;Reference number: A75001
 A;Accession: B75150
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1177 <KAW>
 A;Cross-references: UNIPROT:Q9V1R8; UNIPARC:UPI00000346A8; GB:AJ248284; GB:AL096836; NID
 A;Experimental source: strain Orsay
 C;Genetics:
 A;Gene: PAB2109
 C;Superfamily: chromosome segregation protein SMC1

Query Match 50.0%; Score 43; DB 2; Length 1177;
 Best Local Similarity 80.0%; Pred. No. 55;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 5 YYLNGRTATR 14
 |:||||| |||
 DB 114 YWLNRRRTATR 123

RESULT 11
 S77524
 chromosome segregation protein smc1 - Synechocystis sp. (strain PCC 6803)
 N;Alternate names: protein sll1120
 C;Species: Synechocystis sp.
 A;Variety: PCC 6803
 C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C;Accession: S77524
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 s.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S77524
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-1200 <KAN>
 A;Cross-references: UNIPROT:P73340; UNIPARC:UPI000003451; EMBL:D90905; GB:AB001339; NID
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Superfamily: chromosome segregation protein SMC1

Query Match 50.0%; Score 43; DB 2; Length 1200;
 Best Local Similarity 70.0%; Pred. No. 56;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 4 HYYLNGRTATR 13
 :||| ||||

Db 130 NYINGETAT 139

RESULT 12

B90656
 IcmP-like protein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
 C;Accession: B90656
 R;Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gaeawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: B90656
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1035 <HAY>
 A;Cross-references: UNIPROT:Q8X7W9; UNIPARC:UPI00001653A6; GB:BA000007; PIDN:BA033641.1;
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs0218

Query Match 49.4%; Score 42.5; DB 2; Length 1035;
 Best Local Similarity 87.5%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 WHHYVLNG 9
 ||||| |||
 DB 329 WHHY-NG 335

RESULT 13

C85507
 probable macrophage toxin [imported] - Escherichia coli (strain O157:H7, substrain EDL933;
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
 C;Accession: C85507
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: C85507
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1144 <STO>
 A;Cross-references: UNIPROT:Q8X7W9; UNIPARC:UPI000000D09DE; GB:AE005174; NID:G12512953; P
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z0250

Query Match 49.4%; Score 42.5; DB 2; Length 1144;
 Best Local Similarity 87.5%; Pred. No. 64;
 Matches 7; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 WHHYVLNG 9
 ||||| |||
 DB 438 WHHY-NG 444

RESULT 14

S36824
 beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 22-Jan-1994 #sequence_revision 13-Mar-1997 #text_change 05-Oct-2004
 C;Accession: S36824
 R;Lee, Y.C.; Kurosawa, N.; Hamamoto, T.; Nakaoka, T.; Tsuji, S.
 Eur. J. Biochem. 216, 377-385, 1993
 A;Title: Molecular cloning and expression of Gal-beta-1,3galNAc-alpha-2,3-sialyltransferase
 A;Reference number: S36824; MUID:93387288; PMID:8375377
 A;Accession: S36824
 A;Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-337 <LEE>
A:Cross-references: UNIPROT:P54751; UNIPARC:UPI0000003FB1; EMBL:X73523; NID:g402214; PID
C:Superfamily: sialyltransferase
C:Keywords: glycosyltransferase

Query Match 48.8%; Score 42; DB 2; Length 337;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 WHYYLNGRTA 12
||||:|:|
Db 294 WHYYWNPSPA 304

RESULT 15
I54229
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Oct-2004
C:Accession: I54229; A54898
R:Chang, M.L.; Eddy, R.L.; Shows, T.B.; Lau, J.T.
Glycobiology 5, 319-325, 1995
A:Title: Three genes that encode human beta-galactoside alpha 2,3-sialyltransferases. S
A:Reference number: I54229; MUID:95383839; PMID:7655169
A:Accession: I54229
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-340 <RES>
A:Cross-references: UNIPROT:Q11201; UNIPARC:UPI00000015E1; GB:L13972; NID:g410225; PIDN:
R:Kitagawa, H.; Paulson, J.C.
J. Biol. Chem. 269, 17872-17878, 1994
A:Title: Differential expression of five sialyltransferase genes in human tissues.
A:Reference number: A54898; MUID:94299495; PMID:8027041
A:Accession: A54898
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-11,'V',13-340 <KIT>
A:Cross-references: UNIPARC:UPI000014905A; GB:L29555; NID:g522196; PIDN:AAA36612.1; PID:
C:Genetics:
A:Gene: GDB:STAT4A
A:Cross-references: GDB:384704
A:Map position: 3q21-3q28
C:Superfamily: sialyltransferase
C:Keywords: glycosyltransferase

Query Match 48.8%; Score 42; DB 2; Length 340;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 WHYYLNGRTA 12
||||:|:|
Db 297 WHYYWNPSPA 307

RESULT 16
A45073
Gal beta 1,3GalNAc alpha 2,3-sialyltransferase - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Oct-2004
C:Accession: A45073
R:Gillespie, W.; Kelm, S.; Paulson, J.C.
J. Biol. Chem. 267, 21004-21010, 1992
A:Title: Cloning and expression of the Gal beta 1, 3GalNAc alpha 2,3-sialyltransferase.
A:Reference number: A45073; MUID:93016016; PMID:1383214
A:Accession: A45073
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-343 <GIL>
A:Cross-references: UNIPARC:UPI0000149053
A:Note: sequence extracted from NCBI backbone (NCBIN:116168, NCBI:P:116169)
C:Superfamily: sialyltransferase

Query Match 48.8%; Score 42; DB 2; Length 343;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 WHYYLNGRTA 12
||||:|:|
Db 300 WHYYWNPSPA 310

RESULT 17
AD3012
conserved hypothetical protein Atu3704 [imported] - Agrobacterium tumefaciens (strain C58
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AD3012
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-396 <KUR>
A:Cross-references: UNIPROT:Q8U9M4; UNIPARC:UPI0000164859; GB:AE008689; PIDN:AAL44514.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3704
A:Map position: linear chromosome

Query Match 48.8%; Score 42; DB 2; Length 396;
Best Local Similarity 52.9%; Pred. No. 25;
Matches 9; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

Qy 2 WH-----HYLLNGRTA 12
|||:|:|:|
Db 339 WHVAIVVAVVYLDGRTA 355

RESULT 18
C98272
hypothetical protein AGR_L_2267 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: C98272
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, B.
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: C98272
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-492 <KUR>
A:Cross-references: UNIPROT:Q8U9M4; UNIPARC:UPI000000D22D7; GB:AE007870; PIDN:AAK89701.1;
C:Genetics:
A:Gene: AGR_L_2267
A:Map position: linear chromosome

Query Match 48.8%; Score 42; DB 2; Length 492;
Best Local Similarity 52.9%; Pred. No. 31;
Matches 9; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

Qy 2 WH-----HYLLNGRTA 12
|||:|:|:|
Db 435 WHVAIVVAVVYLDGRTA 451

RESULT 19
A46054

GTP-binding protein ARD 1 - human
N:Alternate names: ADP-ribosylation factor homolog
C:Species: Homo sapiens (man)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Feb-2001
C:Accession: A46054
R:Mishima, K.; Tsuchiya, M.; Nightingale, M.S.; Moss, J.; Vaughan, M.
J. Biol. Chem. 268, 8801-8807, 1993
A:Title: ARD 1, a 64-kDa guanine nucleotide-binding protein with a carboxyl-terminal ADP
A:Reference number: A46054; MUID:93232038; PMID:8473324
A:Accession: A46054
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-574 <MIS>
A:Cross-references: UNIPARC:UPI000017C149
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:129620, NCBIPI:129622)
C:Genetics:
A:Gene: GDB:ARD1; ARPD1
A:Cross-references: GDB:139213; OMIM:601747
A:Map position: 11q23.3-11q23.3
C:Keywords: GTP binding; nucleotide binding; P-loop
F:411-418/Region: nucleotide-binding motif A (P-loop)
F:513-516/Region: GTP-binding NKXD motif

Query Match 48.8%; Score 42; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLYN 8
| | | | |
Db 465 WKHYLYN 471

RESULT 20
S67483
adenosinetriphosphatase 2 - malaria parasite (Plasmodium falciparum)
N:Alternate names: ATPase 2
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
C:Accession: S67483
R:Trottein, F.; Cowman, A.F.
Eur. J. Biochem. 227, 214-225, 1995
A:Title: Molecular cloning and sequence of two novel P-type adenosinetriphosphatases frc
A:Reference number: S67483; MUID:95154293; PMID:7851389
A:Accession: S67483
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1553 <TRO>
A:Cross-references: UNIPROT:Q9U421; UNIPARC:UPI000017B5ED; EMBL:U16955
C:Genetics:
A:Introns: 17/2

Query Match 48.8%; Score 42; DB 2; Length 1553;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHYYLNGRT 11
| | | | |
Db 1345 HHYYFNIRT 1353

RESULT 21
A61259
glycoprotein S - porcine transmissible gastroenteritis virus (strain Miller) (fragments)
N:Alternate names: E2 glycoprotein
C:Species: porcine transmissible gastroenteritis virus
C:Date: 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 09-Jul-2004
C:Accession: A61259; A33140; E33140
R:Bae, I.; Jackwood, D.J.; Benfield, D.A.; Saif, L.J.; Wesley, R.D.; Hill, H.
J. Clin. Microbiol. 29, 215-218, 1991
A:Title: Differentiation of transmissible gastroenteritis virus from porcine respiratory
he S glycoprotein gene.
A:Reference number: A61259; MUID:91131785; PMID:1847152

A:Accession: A61259
A:Status: not compared with conceptual translation
A:Molecule type: genomic RNA
A:Residues: 1-216 <BAE>
A:Cross-references: UNIPROT:Q7LZU7; UNIPARC:UPI0000178664
C:Superfamily: coronavirus E2 glycoprotein
C:Keywords: glycoprotein

Query Match 47.7%; Score 41; DB 2; Length 216;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLYN 9
| | | | |
Db 192 WGHFYNG 199

RESULT 22
AB2349
polysaccharide deacetylase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 31-Dec-2004
C:Accession: AB2349
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2349
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <KUR>
A:Cross-references: UNIPROT:Q8YP55; UNIPARC:UPI000000CEACE; GB:BA0000019; PIDN:BAB76044.1;
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all4345
C:Superfamily: nodulation protein nodB

Query Match 47.7%; Score 41; DB 2; Length 305;
Best Local Similarity 83.3%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6
| | | | |
Db 165 TWHHWY 170

RESULT 23
G82577
phage-related integrase XF2288 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: G82577
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-353 <SIM>
A:Cross-references: UNIPROT:Q9PB56; UNIPARC:UPI000000C29CA; GB:AE0004040; GB:AE0003849; NID:
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
A:Neto, E.; Docena, C.; El-Dorri, H.; Pacinani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A.; Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XP2288

Query Match 47.7%; Score 41; DB 2; Length 353;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHHYLNGRT 11
| : |||||
Db 27 WESFYNGRT 36

RESULT 24
T41390
zinc finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-Oct-2004
C:Accession: T41390
R;Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21390
A:Accession: T41390
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-463 <MUR>
A:Cross-references: UNIPROT:O59811; UNIPARC:UPI000006C562; EMBL:AL023592; PIDN:CAA19119.
A:Experimental source: strain 972h-; cosmid c550
C:Genetics:
A:Gene: SPDB:SPCC550.15C
A:Map position: 3
C:Superfamily: zinc finger protein

Query Match 47.7%; Score 41; DB 2; Length 463;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHHYLNGRTAT 13
| : ||||| : :
Db 28 WHHYLNRKVAS 39

RESULT 25
E81182
hypothetical protein NMB0570 [imported] - Neisseria meningitidis (strain MC58 serogroup
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: E81182
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: E81182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-473 <TET>
A:Cross-references: UNIPROT:Q9K0M2; UNIPARC:UPI00000C44F5; GB:AE002413; GB:AE002098; NID
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0570

Query Match 47.7%; Score 41; DB 2; Length 473;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6
| : |||||
Db 51 TWHHY 56

RESULT 26
B81919
probable membrane protein NMA0753 [imported] - Neisseria meningitidis (strain Z2491 sero
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C:Accession: B81919
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: B81919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-473 <PAR>
A:Cross-references: UNIPROT:Q9JVP7; UNIPARC:UPI00000C4A32; GB:AL162754; GB:AL157959; NID:
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0753

Query Match 47.7%; Score 41; DB 2; Length 473;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6
| : |||||
Db 51 TWHHY 56

RESULT 27
S24284
E2 glycoprotein precursor - porcine respiratory virus (strain 86/137004)
N:Alternate names: peplomer glycoprotein; spike glycoprotein
C:Species: porcine respiratory virus
A:Variety: strain 86/137004
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C:Accession: S24284; S21871
R;Britton, P.; Mawditt, K.L.; Page, K.W.
Virus Res. 21, 181-198, 1991
A:Title: The cloning and sequencing of the virion protein genes from a British isolate of
A:Reference number: S24279; MUID:92116634; PMID:1662846
A:Accession: S24284
A:Molecule type: genomic RNA
A:Residues: 1-1225 <BRI>
A:Cross-references: UNIPROT:P27655; UNIPARC:UPI0000138673; EMBL:X60089; NID:G58983; PIDN:
A:Experimental source: strain 86/137004
C:Superfamily: coronavirus E2 glycoprotein
C:Keywords: Glycoprotein; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-1225/Product: E2 glycoprotein #status predicted <B2G>
F:1167-1187/Domain: transmembrane #status predicted <TMN>
F:26,61,110,121,138,151,181,225,256,292,308,330,338,370,480,501,556,595,610,616,697,850,9
Query Match 47.7%; Score 41; DB 1; Length 1225;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLANG 9
| : |||||
Db 203 WGHFYING 210

RESULT 28
A36607
E2 glycoprotein - porcine respiratory virus (strain RM4)
C:Species: porcine respiratory virus
C>Date: 30-Aug-1991 #sequence_revision 30-Aug-1991 #text_change 09-Jul-2004
C:Accession: A36607

R;Rasschaert, D.; Duarte, M.; Laude, H.
J. Gen. Virol. 71, 2599-2607, 1990
A;Title: Porcine respiratory coronavirus differs from transmissible gastroenteritis virus
A;Reference number: A36607; MUID:91073120; PMID:2174956
A;Accession: A36607
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1225 <RAS>
A;Cross-references: UNIPROT:P24413; UNIPARC:UPI0000138674; GB:224675; NID:g395057; PIDN:
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein

Query Match 47.7%; Score 41; DB 2; Length 1225;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9
| | : | : |
Db 203 WGHFYNG 210

RESULT 29
VGTHF3
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Purdue)
N;Alternate names: peplomer protein; spike glycoprotein
C;Species: porcine transmissible gastroenteritis virus
C;Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 12-Apr-1996
C;Accession: JS0336
R;Jacobs, L.; de Groot, R.; van der Zeijst, B.A.M.; Horzinek, M.C.; Spaan, W.
Virus Res. 8, 363-371, 1987
A;Title: The nucleotide sequence of the peplomer gene of porcine transmissible gastroent
s (FIPV).
A;Reference number: JS0336; MUID:88129049; PMID:2829461
A;Accession: JS0336
A;Molecule type: mRNA
A;Residues: 1-1447 <JAC>
A;Cross-references: UNIPARC:UPI0000174AGE
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1447/Product: E2 glycoprotein #status predicted <MAT>
F;1387-1431/Domain: transmembrane #status predicted <TMN>
F;26,42,71,94,243,250,285,334,345,362,403,447,514,530,552,592,702,723,778,817,832,838,91

Query Match 47.7%; Score 41; DB 1; Length 1447;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9
| | : | : |
Db 425 WGHFYNG 432

RESULT 30
VGTHF2
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Purdue-1
N;Alternate names: spike glycoprotein
C;Species: porcine transmissible gastroenteritis virus
A;Variety: strain Purdue-115
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C;Accession: A27106; S01738
R;Rasschaert, D.; Laude, H.
J. Gen. Virol. 68, 1883-1890, 1987
A;Title: The predicted primary structure of the peplomer protein E2 of the porcine coron
A;Reference number: A27106; MUID:87253116; PMID:3037011
A;Accession: A27106
A;Molecule type: Genomic RNA
A;Residues: 1-1447 <RAS>
A;Cross-references: UNIPROT:P07946; UNIPARC:UPI0000138672; GB:X05695; GB:D00118; NID:g59
R;Rasschaert, D.; Geiffi, J.; Laude, H.
Biochimie 69, 591-600, 1987
A;Title: Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization

A;Reference number: S01738; MUID:88078100; PMID:2825819
A;Accession: S01738
A;Molecule type: Genomic RNA
A;Residues: 1434-1447 <RAW>
A;Cross-references: UNIPARC:UPI0000174A6D; EMBL:X06371
A;Experimental source: strain Purdue-115
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1447/Product: E2 glycoprotein #status predicted <MAT>
F;1387-1431/Domain: transmembrane #status predicted <TMN>
F;26,42,71,94,243,250,285,334,345,362,403,447,514,530,552,592,702,723,778,817,832,838,91

Query Match 47.7%; Score 41; DB 1; Length 1447;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9
| | : | : |
Db 425 WGHFYNG 432

RESULT 31
A43573
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Miller)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
C;Species: porcine transmissible gastroenteritis virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: A43573
R;Wesley, R.D.
Adv. Exp. Med. Biol. 276, 301-306, 1990
A;Title: Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequ
A;Reference number: A43573; MUID:91353366; PMID:1966416
A;Accession: A43573
A;Molecule type: Genomic RNA
A;Residues: 1-1449 <WES>
A;Cross-references: UNIPROT:P33470; UNIPARC:UPI0000174A6F; GB:S51223; NID:g234109; PIDN:f
A;Note: the authors translated the codon GAA for residue 388 as Cys
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1449/Product: E2 glycoprotein #status predicted <E2G>
F;1027-1043/Region: hydrophobic
F;1391-1411/Domain: transmembrane #status predicted <TMN>
F;26,42,71,94,243,250,285,334,345,362,375,405,449,516,532,554,594,704,725,780,819,834,840

Query Match 47.7%; Score 41; DB 1; Length 1449;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9
| | : | : |
Db 427 WGHFYNG 434

RESULT 32
VGTHF5
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain FS772/70)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
C;Species: porcine transmissible gastroenteritis virus
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 09-Jul-2004
C;Accession: B43489; S11728
R;Britton, P.; Page, K.W.
Virus Res. 18, 71-80, 1990
A;Title: Sequence of the S gene from a virulent British field isolate of transmissible g
A;Reference number: A43489; MUID:91188698; PMID:1964522
A;Accession: B43489
A;Molecule type: mRNA
A;Residues: 1-1449 <BRI>
A;Cross-references: UNIPROT:P18450; UNIPARC:UPI000013866F; GB:X53128; NID:g61377; PIDN:C
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein

F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1449/Product: E2 glycoprotein #status predicted <E2G>
F;1027-1043/Region: hydrophobic
F;1395-1411/Domain: transmembrane #status predicted <TMN>
F;126,42,71,94,243,250,285,334,345,362,375,405,449,516,532,554,594,704,725,780,819,834,84
d

Query Match 47.7%; Score 41; DB 1; Length 1449;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLLNG 9
| | : | : |
Db 427 WGHFYING 434

RESULT 33
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus
N;Alternate names: envelope protein; spike protein
C;Species: porcine transmissible gastroenteritis virus
C;Date: 23-Nov-1994 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S65851; S47423
R;Chen, C.M.; Cavanagh, D.; Britton, P.
Virus Res. 38, 83-89, 1995
A;Title: Cloning and sequencing of a 8.4-kb region from the 3'-end of a Taiwanese viru
A;Reference number: S65850; MUID:96060227; PMID:8546012
A;Accession: S65851
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: genomic RNA
F;1-1449 <CH2>
A;Residues: 1-1449
A;Cross-references: UNIPROT:Q88510; UNIPARC:UPI00000F855B; EMBL:Z35758; NID:G529246; PID
A;Experimental source: Taiwanese field isolate
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C;Genetics:
A;Gene: S
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1449/Product: E2 glycoprotein #status predicted <E2G>

Query Match 47.7%; Score 41; DB 2; Length 1449;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLLNG 9
| | : | : |
Db 427 WGHFYING 434

RESULT 34
E2 glycoprotein precursor - canine coronavirus (strain Inseave-1)
N;Alternate names: spike glycoprotein
C;Species: canine coronavirus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jul-2004
C;Accession: JQ1719
R;Horsburgh, B.C.; Brierley, I.; Brown, T.D.K.
J. Gen. Virol. 73, 2849-2862, 1992
A;Title: Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA
A;Reference number: PQ0481; MUID:93057357; PMID:1431811
A;Accession: JQ1719
A;Molecule type: genomic RNA
A;Residues: 1-1451 <HOR>
A;Cross-references: UNIPROT:P36300; UNIPARC:UPI0000138669; DDBJ:D13096; NID:G406193; PID
C;Genetics:
A;Gene: S
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-1451/Product: spike glycoprotein #status predicted <NAT>
F;1394-1412/Domain: transmembrane #status predicted <TMN>
F;28,66,94,142,175,209,235,242,289,338,349,366,379,409,453,520,536,557,707,728,783,821,8

d

Query Match 47.7%; Score 41; DB 1; Length 1451;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLLNG 9
| | : | : |
Db 431 WGHFYING 438

RESULT 35
VG1H79
E2 glycoprotein precursor - feline infectious peritonitis virus (strain 79-1146)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
C;Species: feline infectious peritonitis virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C;Accession: A27171
R;De Groot, R.J.; Maduro, J.; Lenstra, J.A.; Horzinek, M.C.; Van Der Zeijst, B.A.M.; Spa
J. Gen. Virol. 68, 2639-2646, 1987
A;Title: cDNA cloning and sequence analysis of the gene encoding the peplomer protein of
A;Reference number: A27171; MUID:88034948; PMID:3312491
A;Accession: A27171
A;Molecule type: genomic RNA
A;Residues: 1-1452 <DEG>
A;Cross-references: UNIPROT:P10033; UNIPARC:UPI000005FIAB; GB:X06170; GB:D00150; NID:G58;
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;1-19/Domain: transmembrane #status predicted <TM1>
F;20-1452/Product: spike glycoprotein #status predicted <SPG>
F;1394-1414/Domain: transmembrane #status predicted <TM2>
F;29,95,174,208,234,241,288,337,348,365,408,452,483,519,535,557,565,707,728,783,822,837,8
ted

Query Match 47.7%; Score 41; DB 1; Length 1452;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLLNG 9
| | : | : |
Db 430 WGHFYING 437

RESULT 36
S41453
spike protein - canine coronavirus
C;Species: canine coronavirus
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S41453
R;Wesseling, J.G.; Vennema, H.; Godeke, G.J.; Spaan, W.J.M.; Horzinek, M.C.; Rottier, P.
submitted to the EMBL Data Library, December 1993
A;Description: Nucleotide sequence and expression of the spike (S) gene of canine corona
A;Reference number: S41453
A;Accession: S41453
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1453 <WES>
A;Cross-references: UNIPROT:Q65984; UNIPARC:UPI00000F580F; EMBL:X77047; NID:G452379; PID
C;Superfamily: coronavirus E2 glycoprotein

Query Match 47.7%; Score 41; DB 2; Length 1453;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHHYLLNG 9
| | : | : |
Db 431 WGHFYING 438

RESULT 37
AG2106
hypothetical protein alr2406 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AG2106
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2106
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <KUR>
A;Cross-references: UNIPROT:Q44510; UNIPARC:UPI000013BBED; GB:BA000019; PIDN:BA074105.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr2406

Query Match 47.1%; Score 40.5; DB 2; Length 250;
Best Local Similarity 57.1%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 TWH-HYYLNGRTAT 13
||| |||| : |
Db 104 TWHHYYLQRLFT 117

RESULT 38
B88088
protein B0454.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: B88088
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999.
A;Accession: B88088
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-324 <STO>
A;Cross-references: UNIPROT:O17172; UNIPARC:UPI00000797C5; GB:chr_II; PIDN:AB070940.1; PIDN:AB070940.1; PIDN:AB070940.1;
C;Genetics:
A;Gene: B0454.2
A;Map position: 2

A;Residues: 1-74 <TET>
A;Cross-references: UNIPROT:Q9K0F7; UNIPARC:UPI000004460F; GB:AE002420; GB:AE002098; NID:
A;Experimental source: serogroup B, strain MCS8
C;Genetics:
A;Gene: NMB0649

Query	1 TWHHYYLNGR 10	Matches	6; Conservative	1; Mismatches	3; Indels	0; Gaps	0;
Db	41 TWHHQDTGR 50	:					

Query Match 46.5%; Score 40; DB 2; Length 74;
Best Local Similarity 60.0%; Pred. No. 8.9;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

RESULT 40
C81931
Hypothetical protein NMA0855 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C;Accession: C81931
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: C81931
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-81 <PAR>
A;Cross-references: UNIPROT:Q9JVF9; UNIPARC:UPI00000C4A77; GB:AL162754; GB:AL157959; NID:
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0855

A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biological processes
A:Accession: G88008
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 287, 2022-2023, 1999
A:Reference number: AF5000; MUID:99089613; PMID:98519146
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 287, 2022-2023, 1999
A:Accession: B88008
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-324 <STO>
A:Cross-references: UNIPROT:O17172; UNIPARC:UPI00000797C5; GB:chr_II; PIDN:AB70940.1; PIR:G00000
C:Genetics:
A:Gene: B0454.2
A:Map position: 2

```

Query Match      46.5%; Score 40; DB 2; Length 81;
Best Local Similarity 60.0%; Pred. NO. 9.8;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TWHHYLNGR 10
        ||||: ||
Db      48 TWHHQDTGR 57
        ||||: ||

RESULT 41
T21783
hypothetical protein F352.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21783
R:Lennard, N.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19471
A:Accession: T21783
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-227 <WIL>
A:Cross-references: UNIPROT:O62229; UNIPARC:UPI0000077B3E; EMBL:Z81528; PIDN: CAB04289.1;
A:Experimental source: clone F35E2
C:Genetics:
A:Gene: CESP:F35E2.3
A:Map position: 1
A:Introns: 42/1; 102/3; 175/1

```

```

Query Match          47.1%; Score 40.5; DB 2; Length 324;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 2; Indels 3; Gaps 1;

QY      2 WH---HYVLNGRTA 12
      |||::|
Db       11 WHLIYHYVYISGTIA 24

RESULT 39
D81173
hypothetical protein NMB0649 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: D81173
R;Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: D81173
A;Status: preliminary
A;Molecule type: DNA

```

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21783
R:Lennard, N.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19471
A:Accession: T21783
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-227 <WIL>
A:Cross-references: UNIPROT:O62229; UNIPARC:UPI0000077B3E; EMBL:Z81528; PIDN:CAB04289.1;
A:Experimental source: Clone F35E2
C:Genetics:
A:Gene: CESP:F35E2.3
A:Map position: 1
A:Introns: 42/1; 102/3; 175/1

Query Match 46.5%; Score 40; DB 2; Length 227;
Best Local Similarity 54.5%; Pred. No. 30;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TWHHYLLNGRT 11
|:|:|:|:|
Db 207 TYHHYVDNST 217

```

RESULT 42
AD1209
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) lmo1076 - Listeria monocytogenes (strain
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1209
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <GLA>
A:Cross-references: UNIPROT:Q8Y842; UNIPARC:UPI0000054CB2; GB:NC_003210; PIDN:CAC99154.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo1076
C:Keywords: hydrolase

Query Match 46.5%; Score 40; DB 2; Length 572;
Best Local Similarity 45.5%; Pred. No. 80;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 TWHYYLNGRT 11
||: : |||:
Db 472 TWYQFVNGKT 482

RESULT 43
T08984
auxin response factor 7 homolog F6G3.110 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C:Accession: T08984
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16520
A:Accession: T08984
A:Molecule type: DNA
A:Residues: 1-653 <BEV>
A:Cross-references: UNIPARC:UPI0000A0314; EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.110
A:Experimental source: cultivar Columbia; BAC clone F6G3
C:Genetics:
A:Gene: ATSP:F6G3.110
A:Map position: 4
A:Introns: 362/3

Query Match 46.5%; Score 40; DB 2; Length 653;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHYYLN 8
|||||
Db 464 HHYYLN 469

RESULT 44
T37276
protein PEX6 homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T37276; T03904; T37307
R:Kamiryo, T.; Bun-ya, M.
submitted to the EMBL Data Library, February 1998
A:Description: C.elegans cDNA coding PEX6 homologue.
A:Reference number: Z21663
A:Accession: T37276
A>Status: preliminary; translated from GB/EMBL/DDBJ

```

```

A:Molecule type: mRNA
A:Residues: 1-720 <KAM>
A:Cross-references: UNIPROT:O16270; UNIPARC:UPI000007F9CB; EMBL:AB010968; PIDN:BAA33544.1
R:Du, Z.; Le, T.T.; Holmes, A.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of C. elegans cosmid F39G3.
A:Reference number: Z15131
A:Accession: T03904
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-720 <DUZ>
A:Cross-references: UNIPARC:UPI000007F9CB; EMBL:AF016424; NID:G2291203; PIDN:AAB65332.1.
R:Kamiryo, T.; Bun-ya, M.; Ghenea, S.
submitted to the EMBL Data Library, March 1998
A:Description: Genomic DNA sequence of PEX6 homolog of C. elegans.
A:Reference number: Z21673
A:Accession: T37307
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-720 <KA2>
A:Cross-references: UNIPARC:UPI000007F9CB; EMBL:AB012224; PIDN:BAA76440.1
C:Genetics:
A:Map position: V
A:Introns: 47/2; 122/3; 170/2; 338/1; 458/3; 659/3
A:Note: F39G3.7

Query Match 46.5%; Score 40; DB 2; Length 720;
Best Local Similarity 58.3%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTA 12
||| : |||: |
Db 393 TWLOYLYNEKLA 404

RESULT 45
E97177
uncharacterized conserved membrane protein, affecting LPS biosynthesis CAC2251 [imported]
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 05-Oct-2004
C:Accession: E97177
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97177
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-723 <KUR>
A:Cross-references: UNIPROT:Q97GW4; UNIPARC:UPI00000CA479; GB:AE001437; PIDN:AAK80208.1;
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2251
C:Superfamily: uncharacterized conserved protein

Query Match 46.5%; Score 40; DB 2; Length 723;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYL 7
||| |||
Db 416 TWHPYL 422

RESULT 46
F88448
protein C45G9.10 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F88448
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

```

A>Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biological processes
A/Reference number: AY5000; MUID:99069613; PMID:9851916
A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 35, 1999
A/Accession: F88448
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-971 <STO>
A/Cross-references: UNIPROT:Q09281; UNIPARC:UPI000013BF17; GB:chr_III; PIDN:AAA62553.1; C/Genetics:
A/Gene: C45G9.10
A/Map position: 3

Query Match 46.5%; Score 40; DB 2; Length 971;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHYYLNGRT 11
| | | | |
Db 696 HKYYINGET 704

RESULT 47
S46837
hypothetical protein YHL023c - yeast (*Saccharomyces cerevisiae*)
C/Species: *Saccharomyces cerevisiae*
C/Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C/Accession: S46837
R/Favell, T.
submitted to the EMBL Data Library, June 1994
A/Description: The sequence of *S. cerevisiae* cosmid 9433.
A/Reference number: S46796
A/Accession: S46837
A/Molecule type: DNA
A/Residues: 1-1146 <FAV>
A/Cross-references: UNIPROT:P38742; UNIPARC:UPI000013B1D0; EMBL:U11582; NID:g2289793; PIDN:AAA62553.1
C/Genetics:
A/Gene: MIPS:YHL023c
A/Cross-references: SGD:S0001015
A/Map position: 8L

Query Match 46.5%; Score 40; DB 2; Length 1146;
Best Local Similarity 46.2%; Pred. No. 1.7e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHYYLNGRTAT 14
: | | | | | | | |
Db 343 YHHYHKNATSOR 355

RESULT 48
H82802
fimbrial assembly protein XF0478 [imported] - *Xylella fastidiosa* (strain 9a5c)
C/Species: *Xylella fastidiosa*
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C/Accession: H82802
R/Anonymous, T.
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A/Reference number: H82515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: H82802
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1472 <SIM>
A/Cross-references: UNIPROT:Q96G24; UNIPARC:UPI00000C2413; GB:AE003897; GB:AE003849; NID:g2289793
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z. A/Reference number: A59328
A/Contents: annotation
C/Genetics:
A/Gene: XF0478

Query Match 46.5%; Score 40; DB 2; Length 1472;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHYYLNGR 10
| | | | |
Db 1032 HHYYVDGK 1039

RESULT 49
E83765
hypothetical protein BH0925 [imported] - *Bacillus halodurans* (strain C-125)
C/Species: *Bacillus halodurans*
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: E83765
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-108 <STO>
A/Cross-references: UNIPROT:Q9KEC8; UNIPARC:UPI00000C3A2F; GB:AP001510; GB:BA000004; NID:g2289793
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH0925
C/Superfamily: *Bacillus subtilis* hypothetical protein yfhh

Query Match 45.3%; Score 39; DB 2; Length 108;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTA 12
| : | | | | | | | |
Db 71 TFHISYLNGRFA 82

RESULT 50
KLCHI
calcium-binding protein, vitamin D-dependent - chicken
N/Alternate names: avian type CaBP; calbindin; large CaBP
C/Species: *Gallus gallus* (chicken)
C/Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 09-Jul-2004
C/Accession: A40926; A93605; A94129; A29957; A94166; A03070; A28026; JT0298
R/Minghetti, P.P.; Cancala, L.; Fujisawa, Y.; Theofan, G.; Norman, A.W. Mol. Endocrinol. 2, 355-367, 1988
A/Title: Molecular structure of the chicken vitamin D-induced calbindin-D-28k gene reveal A/Reference number: A40926; MUID:88246474; PMID:2967915
A/Accession: A40926
A/Molecule type: DNA
A/Residues: 1-262 <MIN>
A/Cross-references: UNIPROT:P04354; UNIPARC:UPI00001712BA; GB:M31139
R/Wilson, P.W.; Harding, M.; Lawson, D.E.M.
Nucleic Acids Res. 13, 8867-8881, 1985
A/Title: Putative amino acid sequence of chick calcium-binding protein deduced from a con A/Reference number: A93605; MUID:86093684; PMID:3841205
A/Accession: A93605
A/Molecule type: mRNA
A/Residues: 1-262 <WTL>
A/Cross-references: UNIPARC:UPI00001712BA; GB:X03343; NID:g63170; PIDN:CAA27049.1; PID:g63170
A/Note: the authors translated the codon ACU for residue 70 as Tyr and GAU for residues 1

R;Hunziker, W.
 Proc. Natl. Acad. Sci. U.S.A. 83, 7578-7582, 1986
 A;Title: The 28-kDa vitamin D-dependent calcium-binding protein has a six-domain structure
 A;Reference number: A94129; MUID:87016992; PMID:3463988
 A;Accession: A94129
 A:Molecule type: mRNA
 A:Residues: 1-262 <HUN>
 A:CROSS-references: UNIPARC:UPI00001712BA; GB:W14230; NID:G211429; PIDN:AAA48659.1; PID:
 R;Wilson, P.W.; Rogers, J.; Harding, M.; Pohl, V.; Pattyn, G.; Lawson, D.E.M.
 J. Mol. Biol. 200, 615-625, 1998
 A;Title: Structure of chick chromosomal genes for calbindin and calretinin.
 A;Reference number: A29957; MUID:88316929; PMID:3411606
 A;Accession: A29957
 A:Molecule type: DNA
 A:Residues: 1-262 <WI2>
 A:CROSS-references: UNIPARC:UPI00001712BA; EMBL:X06629; NID:G63133; PIDN:CAA29843.1; PID:
 R;Fullmer, C.S.; Wasserman, R.H.
 Proc. Natl. Acad. Sci. U.S.A. 84, 4772-4776, 1987
 A;Title: Chicken intestinal 28-kilodalton calbindin-D: complete amino acid sequence and
 A;Reference number: A94166; MUID:87260872; PMID:3474624
 A;Accession: A94166
 A:Molecule type: protein
 A:Residues: 2-262 <FUL>
 A:CROSS-references: UNIPARC:UPI0000126D72
 C;Comment: This protein is involved in vitamin D-stimulated calcium absorption from the
 D; without it, synthesis quickly stops.
 C;Comment: Four calcium ions can be bound per molecule.
 C;Comment: The six domains of this protein are thought to derive from duplication of an
 -86 and 231-260, respectively) no longer bind calcium.
 C;Genetics:
 A;Introns: 28/1; 53/3; 78/3; 106/3; 125/3; 151/3; 170/2; 183/3; 201/3; 225/3
 C;Superfamily: calretinin; calmodulin repeat homology
 C;Keywords: blocked amino end; calcium binding; duplication; EF hand; vitamin D
 F;2-262/Product: calcium-binding protein, vitamin D-dependent #status experimental <MAT>
 F;12-44/Domain: calmodulin repeat homology <EF1>
 F;54-86/Domain: calmodulin repeat homology <EF2>
 F;99-131/Domain: calmodulin repeat homology <EF3>
 F;143-175/Domain: calmodulin repeat homology <EF4>
 F;187-219/Domain: calmodulin repeat homology <EF5>
 F;2/Modified site: blocked amino end (Thr) (in mature form) #status experimental
 F;25-27,29,31,36/Binding site: calcium (Asp, Asn, Tyr, Glu) #status predicted
 F;112,114,116,118,123/Binding site: calcium (Asp, Asp, Ser, Phe, Glu) #status predicted
 F;156,158,160,162,167/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted
 F;200,202,204,206,211/Binding site: calcium (Asp, Asp, Asn, Tyr, Glu) #status predicted

Query Match 45.3%; Score 39; DB 1; Length 262;
 Best Local Similarity 40.0%; Pred.No. 51;
 Matches 6; Conservative 3; Mismatches 0; Indels 6; Gaps 1;

QY 2 WHHY-----YLNGR 10
 |||||
 |||||
 Db 21 WHHYDSGNGYMDGK 35

Search completed: June 5, 2006, 12:53:54
 Job time : 25.5342 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 122.74 Seconds
(without alignments)
105.510 Million cell updates/sec

Title: US-10-645-659A-10

Perfect score: 86

Sequence: 1 TWHYYLNGRTATR 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Uniprot 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	100.0	543	1	HPSE_HUMAN
2	79	91.9	523	1	HPSE_CHICK
3	78	90.7	536	1	HPSE_RAT
4	77	89.5	535	1	HPSE_MOUSE
5	77	89.5	574	2	Q333X6_SPAJD
6	77	89.5	574	2	Q333X7_GRODE
7	77	89.5	574	2	Q333X8_GRODE
8	77	89.5	574	2	Q333X9_GRODE
9	74	86.0	545	1	HPSE_BOVIN
10	66	76.7	255	2	Q4TGC8_TETNG
11	66	76.7	533	2	Q4SVR6_TETNG
12	55	64.0	515	2	Q8T108_BOMMO
13	52	60.5	532	2	Q9PTB2_ICTPU
14	52	60.5	597	2	Q4TB80_TETNG
15	51	59.3	1179	2	Q59462_PYRHO
16	50	58.1	124	2	Q4MRV8_ASPFU
17	49	57.0	52	2	Q2KXK8_BORAV
18	47	54.7	138	2	Q7P1C9_CHRVO
19	47	54.7	215	2	Q5C4P3_SCHJA
20	47	54.7	382	2	Q7VJ45_HELHP
21	47	54.7	383	2	Q41029_GIBZE
22	47	54.7	1159	1	KCMA1_CANFA
23	46	53.5	64	2	Q8BSE9_MOUSE
24	46	53.5	186	2	Q418T2_GIBZE
25	46	53.5	190	2	Q86ZK9_PODAN
26	46	53.5	208	2	Q2UL82_ASPOR
27	46	53.5	213	2	Q7S4E2_NEUCR
28	46	53.5	213	2	Q5B387_EMENI
29	46	53.5	279	2	Q4WX16_ASPFU
30	46	53.5	301	2	Q5DZ88_VIBF1
31	46	53.5	349	2	Q8BSA0_MOUSE

32	46	53.5	349	2	Q5ZJJ0_CHICK
33	46	53.5	349	2	Q70D58_CHICK
34	46	53.5	350	1	SLA4B_HUMAN
35	46	53.5	350	1	SLA4B_PANTR
36	46	53.5	350	1	SLA4B_RAT
37	46	53.5	350	2	Q6H8M9_BOVIN
38	46	53.5	350	2	Q8BPL0_MOUSE
39	46	53.5	350	2	Q91WH6_MOUSE
40	46	53.5	351	2	Q6NU41_XENLA
41	46	53.5	892	2	Q4XU3_9BURK
42	45	52.3	190	2	Q80438_ARATH
43	45	52.3	448	2	Q2URZ0_ASPOR
44	45	52.3	558	2	Q333X5_SPAJD
45	45	52.3	612	2	Q4RHH9_TETNG
46	44.5	51.7	3050	2	Q6PY34_9POTY
47	44	51.2	161	2	Q7V6T4_PROMM
48	44	51.2	234	2	Q4SB93_TETNG
49	44	51.2	236	2	Q404U7_9ROB
50	44	51.2	260	2	Q5W6T7_ORYSA
51	44	51.2	290	2	Q8YW23_ANASP
52	44	51.2	470	2	Q47YS4_COLP3
53	44	51.2	492	2	Q6PGT9_BRARE
54	44	51.2	522	2	Q9LIW0_ORYSA
55	44	51.2	1484	2	Q4Q4K3_LEIMA
56	44	51.2	2835	2	Q8G9Q2_LEUME
57	44	51.2	3972	2	Q9S0R8_STRAW
58	44	51.2	5532	2	Q9S0R4_STRAW
59	43.5	50.6	243	2	Q4E770_9RICK
60	43.5	50.6	357	2	Q4ECX5_9RICK
61	43	50.0	181	2	Q7NTK4_CHRVO
62	43	50.0	225	2	Q2WRJ4_CLOBE
63	43	50.0	277	2	Q3M3P2_ANAVT
64	43	50.0	342	1	SLA4A_CHICK
65	43	50.0	375	2	Q8A7Y3_BACTN
66	43	50.0	443	2	Q8A2C5_BACTN
67	43	50.0	479	2	Q49446_ARATH
68	43	50.0	521	2	Q5TMS4_ANOGA
69	43	50.0	556	2	Q7NWL5_CHRVO
70	43	50.0	562	2	Q5PP23_ARATH
71	43	50.0	581	2	Q6A3Q7_9BACT
72	43	50.0	592	1	HPSE2_HUMAN
73	43	50.0	592	2	Q2M1H9_HUMAN
74	43	50.0	608	2	Q5LH86_BACFN
75	43	50.0	608	2	Q4LYS8_BACFR
76	43	50.0	624	2	Q2USY5_ASPOR
77	43	50.0	679	2	Q8A6U5_BACTN
78	43	50.0	690	2	Q2UTE2_ASPOR
79	43	50.0	782	2	Q5SR54_HUMAN
80	43	50.0	857	2	Q4WN22_ASPFU
81	43	50.0	926	2	Q4RVF3_TETNG
82	43	50.0	948	2	Q9DDD3_CHICK
83	43	50.0	971	2	Q5UE58_HUMAN
84	43	50.0	971	2	Q71MN0_HUMAN
85	43	50.0	971	2	Q8N4K9_HUMAN
86	43	50.0	978	1	CSTN1_DROME
87	43	50.0	981	1	CSTN1_HUMAN
88	43	50.0	981	2	Q5SR52_HUMAN
89	43	50.0	1177	2	Q87711_PYRFU
90	43	50.0	1177	2	Q3V1B8_PYRAB
91	43	50.0	1200	2	Q73340_SYNY3
92	43	50.0	1291	2	Q8TZY2_PYRFU
93	43	50.0	1348	2	Q7PZL0_ANOGA
94	43	50.0	1675	2	Q60EN4_ORYSA
95	42.5	49.4	383	2	Q8C3K0_MOUSE
96	42.5	49.4	559	2	Q3Z5D9_SHISS
97	42.5	49.4	894	2	Q2Y774_NITMU
98	42.5	49.4	1035	2	Q7AHJ7_ECO57
99	42.5	49.4	1144	2	Q8X7W9_ECO57
100	42.5	49.4	1258	2	Q80S27_MIDDV
101	42	48.8	41	2	Q54QV6_DICDI
102	42	48.8	94	2	Q8XM53_CLOPE
103	42	48.8	129	2	Q4D9U2_TRYCR
104	42	48.8	146	2	Q7R4L4_GIALA

Q5ZJJ0	gallus gall
Q70D58	gallus gall
Q16842	h cmp-n-ace
Q6K588	p cmp-n-ace
Q11205	r cmp-n-ace
Q6H8M9	bos taurus
Q8BPL0	m o day neo
Q91WH6	mus musculus
Q6NU41	xenopus lae
Q4XU3	polaronomas
Q80438	arabidopsis
Q2URZ0	aspergillus
Q333X5	spalax juda
Q4RHH9	tetraodon n
Q6PY34	hordeum mos
Q7V6T4	prochloroco
Q4SB93	tetraodon n
Q404U7	jannaschia
Q5W6T7	oryza sativ
Q8YW23	anabaena sp
Q47YS4	colwellia p
Q6PGT9	brachydanio
Q9LIW0	oryza sativ
Q4Q4K3	leishmania
Q8G9Q2	leuconostoc
Q9S0R8	streptomyce
Q9S0R4	streptomyce
Q4E770	wolbachia e
Q4ECX5	wolbachia e
Q7NTK4	chromobacte
Q2WRJ4	clostridium
Q3M3P2	anabaena va
Q11200	g cmp-n-ace
Q8A7Y3	bacteroides
Q8A2C5	bacteroides
Q49446	arabidopsis
Q5TMS4	anopheles g
Q7NWL5	chromobacte
Q5PP23	arabidopsis
Q6A3Q7	uncultured
Q8WWQ2	homo sapien
Q2M1H9	homo sapien
Q5LH86	bacteroides
Q4LYS8	bacteroides
Q2USY5	aspergillus
Q8A6U5	bacteroides
Q2UTE2	aspergillus
Q5SR54	homo sapien
Q4WN22	aspergillus
Q4RVF3	tetraodon n
Q9DDD3	gallus gall
Q5UE58	homo sapien
Q71MN0	homo sapien
Q8N4K9	homo sapien
Q9V498	drosophila
Q5SR52	homo sapien
Q87711	pyrococcus
Q3V1B8	pyrococcus
Q73340	synecocyst
Q8TZY2	pyrococcus
Q7PZL0	anopheles g
Q60EN4	oryza sativ
Q8C3K0	mus musculus
Q3Z5D9	shigella so
Q2Y774	nitrospir
Q7AHJ7	escherichia
Q8X7W9	escherichia
Q80S27	middelburg
Q54QV6	dictyosteli
Q8XM53	clostridium
Q4D9U2	trypanosoma
Q7R4L4	giardia lam

105	180	2	Q8BJX5_MOUSE	178	41.5	48.3	192	2	O55932_9H1V1	O55932	human immun
106	42	48.8	Q31QA9_SYNBP7	179	41.5	48.3	192	2	O55933_9H1V1	O55933	human immun
107	42	48.8	Q5N3X7_SYNP6	180	41.5	48.3	192	2	O55935_9H1V1	O55935	human immun
108	42	48.8	Q357P8_9BRAD	181	41.5	48.3	192	2	O55936_9H1V1	O55936	human immun
109	42	48.8	Q32S45_9MO1L	182	41.5	48.3	192	2	O55938_9H1V1	O55938	human immun
110	42	48.8	Q5JP11_HUMAN	183	41.5	48.3	192	2	O55941_9H1V1	O55941	human immun
111	42	48.8	Q8DWJ3_STRWD	184	41.5	48.3	192	2	O55942_9H1V1	O55942	human immun
112	42	48.8	Q4WJ49_ASPFPU	185	41.5	48.3	192	2	O55943_9H1V1	O55943	human immun
113	42	48.8	Q7Q2C2_ANOGA	186	41.5	48.3	192	2	O55944_9H1V1	O55944	human immun
114	42	48.8	Q5STI0_HUMAN	187	41.5	48.3	192	2	O55945_9H1V1	O55945	human immun
115	42	48.8	Q5SSW3_HUMAN	188	41.5	48.3	192	2	Q7ZJC5_9H1V1	O7ZJC5	human immun
116	42	48.8	Q33IU9_9CAUD	189	41.5	48.3	442	2	Q4QF00_DESAC	Q4QF00	desulfurom
117	42	48.8	Q9BEG4_BOVIN	190	41.5	48.3	688	1	EFQ_APPPP	EFQ_APPPP	apple proli
118	42	48.8	Q3IU61_MOUSE	191	41.5	48.3	688	2	O6C9N5_YARLI	O6C9N5	yarrowia li
119	42	48.8	Q3UU61_MOUSE	192	41	47.7	66	2	O8QRN4_9CORO	O8QRN4	canine coro
120	42	48.8	Q544T4_MOUSE	193	41	47.7	95	2	O6LXB0_METMP	O6LXB0	methanococ
121	42	48.8	Q33IU9_9CAUD	194	41	47.7	150	2	Q3R9O6_XYLFA	Q3R9O6	xylella fas
122	42	48.8	Q33IU9_9CAUD	195	41	47.7	150	2	Q3R193_XYLFA	Q3R193	xylella fas
123	42	48.8	Q33IU9_9CAUD	196	41	47.7	182	2	Q7UU75_RHOBA	Q7UU75	rhodopirell
124	42	48.8	Q6H8N0_RAT	197	41	47.7	183	2	O89ID8_BRAJA	O89ID8	bradyrhizob
125	42	48.8	Q344A_FIG	198	41	47.7	216	2	Q3PWF9_NITHA	Q3PWF9	nitrobacter
126	42	48.8	Q62WX9_BACID	199	41	47.7	216	2	O7LZU7_9CORO	O7LZU7	transmissib
127	42	48.8	Q65LI8_BACID	200	41	47.7	218	2	Q2VYJ3_MAGSA	Q2VYJ3	magnetospir
128	42	48.8	Q5SSW4_HUMAN	201	41	47.7	226	2	Q9HS16_HUMAN	Q9HS16	homo sapien
129	42	48.8	Q6H8N0_RAT	202	41	47.7	254	2	Q2L2P0_BORAV	Q2L2P0	bordetella
130	42	48.8	Q7MB15_VIBVY	203	41	47.7	260	2	Q2NRG5_SODGL	Q2NRG5	sodalis glo
131	42	48.8	Q8U9M4_AGR75	204	41	47.7	289	2	Q2R2T1_ORYSA	Q2R2T1	oryza sativ
132	42	48.8	Q8C5S6_MOUSE	205	41	47.7	305	2	Q3MDL7_ANAVT	Q3MDL7	anabaena va
133	42	48.8	Q35QK5_9BRAD	206	41	47.7	305	2	O8YP55_ANASP	O8YP55	anabaena sp
134	42	48.8	Q25PH1_MAIZE	207	41	47.7	307	2	O8UZE9_PYRFU	O8UZE9	pyrococcus
135	42	48.8	Q35QK5_9BRAD	208	41	47.7	318	2	O4RHC0_ENTHI	O4RHC0	entamoeba h
136	42	48.8	Q93XP6_MAIZE	209	41	47.7	320	2	O4RHB6_TETNG	O4RHB6	tetraodon n
137	42	48.8	Q7CT44_AGR75	210	41	47.7	322	2	O9U3M4_CAEEL	O9U3M4	caenorhabdi
138	42	48.8	Q5WD67_BACSK	211	41	47.7	331	2	Q2R2S3_ORYSA	Q2R2S3	oryza sativ
139	42	48.8	Q5SSX7_HUMAN	212	41	47.7	333	2	Q3R0X1_XYLFA	Q3R0X1	xylella fas
140	42	48.8	Q5STI1_HUMAN	213	41	47.7	353	2	Q3RSB9_XYLFA	Q3RSB9	xylella fas
141	42	48.8	Q73F21_BACCI	214	41	47.7	353	2	O9PB56_XYLFA	O9PB56	xylella fas
142	42	48.8	Q3GZ29_9ACTO	215	41	47.7	367	2	O2S9G3_9GAMM	O2S9G3	hahella che
143	42	48.8	Q3GZ29_9ACTO	216	41	47.7	374	2	O5NXV2_AZOSE	O5NXV2	azocarcus sp
144	42	48.8	ARD1_RAT	217	41	47.7	388	2	Q3WIQ7_9ACTO	Q3WIQ7	frankia sp
145	42	48.8	ARD1_HUMAN	218	41	47.7	390	2	Q347I8_RHOFA	Q347I8	rhodospseudo
146	42	48.8	ARD1_MOUSE	219	41	47.7	411	2	Q35R75_9BRAD	Q35R75	bradyrhizob
147	42	48.8	Q6DDN3_XENLA	220	41	47.7	413	2	O7G780_ORYSA	O7G780	oryza sativ
148	42	48.8	Q4S258_TETNG	221	41	47.7	428	2	Q02421_ORYSA	Q02421	porcine res
149	42	48.8	Q4WBRA_HORVU	222	41	47.7	428	2	O02422_9CORO	O02422	porcine res
150	42	48.8	RN139_HUMAN	223	41	47.7	428	2	Q02423_9CORO	Q02423	porcine res
151	42	48.8	Q5RBT7_PONPY	224	41	47.7	428	2	Q02424_9CORO	Q02424	porcine res
152	42	48.8	RN139_MOUSE	225	41	47.7	456	2	Q4G033_9CORO	Q4G033	rattus norv
153	42	48.8	Q5Z168_CHICK	226	41	47.7	463	2	O59811_SCHPO	O59811	schizosacch
154	42	48.8	Q9HID7_THEAC	227	41	47.7	473	2	O9JVP7_NEIMA	O9JVP7	neisseria m
155	42	48.8	Q4SKR4_TETNG	228	41	47.7	473	2	O9K0M2_NEIME	O9K0M2	neisseria m
156	42	48.8	Q43KS2_SOLUS	229	41	47.7	476	2	O8MT72_DROME	O8MT72	drosophila
157	42	48.8	CSTN3_HUMAN	230	41	47.7	490	1	SYG_BIFLO	SYG_BIFLO	bifidobacte
158	42	48.8	CSTN3_MOUSE	231	41	47.7	490	2	O5EMH2_9CORO	O5EMH2	canine coro
159	42	48.8	Q2T9J5_HUMAN	232	41	47.7	490	2	O5G1S4_9CORO	O5G1S4	canine coro
160	42	48.8	Q5R9Q9_PONPY	233	41	47.7	497	2	O5F6Y2_NEIG1	O5F6Y2	neisseria g
161	42	48.8	O544R0_MOUSE	234	41	47.7	504	2	Q37W36_SPHAR	Q37W36	novosphingo
162	42	48.8	CSTN3_RAT	235	41	47.7	507	2	Q74MZ6_NANEQ	Q74MZ6	nanosaracha
163	42	48.8	Q5UB57_HUMAN	236	41	47.7	511	1	GPMT_STRCO	GPMT_STRCO	streptococ
164	42	48.8	Q61BN8_CAEER	237	41	47.7	542	2	O5FQW3_GLUOX	O5FQW3	gluconobact
165	42	48.8	Q6TKT8_ECOLI	238	41	47.7	578	2	O6ZUX8_HUMAN	O6ZUX8	homo sapien
166	42	48.8	Q9PUM2_XENLA	239	41	47.7	609	2	O6ZU58_HUMAN	O6ZU58	homo sapien
167	42	48.8	Q27720_PLAFA	240	41	47.7	613	2	O4IB63_GIBZE	O4IB63	gibberella
168	42	48.8	Q815L4_PLAFA	241	41	47.7	621	2	O4SPT3_TETNG	O4SPT3	tetraodon n
169	42	48.8	Q9U421_PLAFA	242	41	47.7	624	2	Q74SQ4_MYCPA	Q74SQ4	mycobacteri
170	41.5	48.3	Q4Q3F1_LEIEM	243	41	47.7	639	2	Q9WIJ9_DROME	Q9WIJ9	drosophila
171	41.5	48.3	Q8QRX7_9BETA	244	41	47.7	668	2	O4P6C8_UBTMA	O4P6C8	ustilago ma
172	41.5	48.3	O55925_9H1V1	245	41	47.7	687	2	Q9DUT2_9BROM	Q9DUT2	pelargonium
173	41.5	48.3	O55926_9H1V1	246	41	47.7	693	2	O5W7F3_BOMMO	O5W7F3	bombyx mori
174	41.5	48.3	O55927_9H1V1	247	41	47.7	753	2	O628L0_CAEER	O628L0	caenorhabdi
175	41.5	48.3	O55928_9H1V1	248	41	47.7	777	2	Q9QRQ1_9CORO	Q9QRQ1	canine coro
176	41.5	48.3	O55929_9H1V1	249	41	47.7	784	2	O65I20_BACILL	O65I20	bacillus li
177	41.5	48.3	O55930_9H1V1	250	41	47.7	789	2	O8QRR7_9CORO	O8QRR7	transmissib

251	41	47.7	789	2	QBQR8_9CORO	QBqrr8	transmissib	324	40	46.5	121	2	Q926M3_LISIN	Q926m3	listeria in
252	41	47.7	789	2	QBQR9_9CORO	QBqrr9	transmissib	325	40	46.5	127	1	ILBP_RABIT	P50119	oryctolagus
253	41	47.7	789	2	QBQR50_9CORO	QBqrs0	transmissib	326	40	46.5	132	2	O22229_CABEL	O62229	caenorhabdi
254	41	47.7	789	2	QBQR53_9CORO	QBqrs3	transmissib	327	40	46.5	137	2	Q9ACRI_STRCO	Q9acri	streptomyc
255	41	47.7	790	2	Q5LIY7_GEOKA	Q5liy7	geobacillus	328	40	46.5	150	2	Q5KUW8_GROKA	Q5kuw8	geobacillus
256	41	47.7	790	2	QBQR51_9CORO	QBqrs1	transmissib	329	40	46.5	155	2	Q8G55_PSEPK	Q8g55	pseudomonas
257	41	47.7	790	2	QBQR52_9CORO	QBqrs2	transmissib	330	40	46.5	182	2	Q5L246_GROKA	Q5l246	geobacillus
258	41	47.7	793	2	QBQR02_9CORO	QBqrq2	canine coro	331	40	46.5	196	2	Q4ZM35_PSEU2	Q4zmj5	pseudomonas
259	41	47.7	808	2	QBQR00_9CORO	QBqrq0	canine coro	332	40	46.5	196	2	Q88A22_PSESM	Q88a22	pseudomonas
260	41	47.7	813	2	QBQR04_9CORO	QBqrq4	canine coro	333	40	46.5	246	2	Q2UBJ1_ASPOR	Q2ubj1	aspergillus
261	41	47.7	815	2	QBQR03_9CORO	QBqrq3	canine coro	334	40	46.5	252	2	Q63LW5_BURPS	Q63lw5	burkholderi
262	41	47.7	883	2	Q55K31_CRYNE	Q55k31	cryptococcu	335	40	46.5	256	2	Q44XV6_9BURK	Q44xv6	burkholderi
263	41	47.7	883	2	Q5K9G8_CRYNE	Q5k9g8	cryptococcu	336	40	46.5	256	2	Q4LV13_9BURK	Q4lv13	burkholderi
264	41	47.7	941	2	Q4P8R4_USTMA	Q4p8r4	ustilago ma	337	40	46.5	256	2	Q3JFR8_BURP1	Q3jfr8	burkholderi
265	41	47.7	1064	2	Q4J7S9_SULAC	Q4j7s9	sulfolobus	338	40	46.5	256	2	Q2T5S3_BURTH	Q2t5s3	burkholderi
266	41	47.7	1167	2	Q5CS41_CRYPV	Q5cs41	cryptospori	339	40	46.5	256	2	Q62BH7_BURMA	Q62bh7	burkholderi
267	41	47.7	1225	1	SPIKE_CVPR8	P27655	porcine res	340	40	46.5	268	2	Q935I7_SALTI	Q935i7	salmonella
268	41	47.7	1225	1	SPIKE_CVPRM	P24413	porcine res	341	40	46.5	274	2	Q8TIQ8_METAC	Q8tiq8	methanosaec
269	41	47.7	1225	2	Q84852_9CORO	Q84852	porcine res	342	40	46.5	282	2	Q3Q562_9GNMM	Q3q562	shewanella
270	41	47.7	1263	2	Q7Z485_HUMAN	Q7z485	homo sapien	343	40	46.5	283	2	Q35G84_9BRAD	Q35g84	bradyrhizob
271	41	47.7	1310	1	AT8B3_HUMAN	O60423	homo sapien	344	40	46.5	284	1	TRUA_SYNEL	Q8cvm6	synecococc
272	41	47.7	1393	2	Q338U3_ORYSA	Q338u3	oryza sativ	345	40	46.5	291	2	Q3WML3_9RHIZ	Q3wwl3	mesorhizobi
273	41	47.7	1429	2	Q5CLH5_CRYHO	O5clh5	cryptospori	346	40	46.5	324	2	Q5LH55_BACFN	Q5lh55	bacteroides
274	41	47.7	1429	2	Q5CV19_CRYPV	O5cv19	cryptospori	347	40	46.5	324	2	Q64YS7_BACFR	Q64ys7	bacteroides
275	41	47.7	1447	1	SPIKE_CVPPU	P07946	porcine tra	348	40	46.5	325	2	Q3GHJ3_CHLVI	Q3ghj3	prosthecoch
276	41	47.7	1447	1	SPIKE_CVPRT	Q1977	porcine tra	349	40	46.5	330	2	Q6KBS7_FUGRU	Q6kbs7	fugu rubrip
277	41	47.7	1447	1	QBQW0_9CORO	QBqww0	transmissib	350	40	46.5	332	2	Q70D56_XENTR	Q70d56	xenopus tro
278	41	47.7	1447	1	Q9IWO4_CRYPO	Q9iwo4	transmissib	351	40	46.5	332	2	Q70D57_XENLA	Q70d57	xenopus lae
279	41	47.7	1449	1	SPIKE_CVPPS	P18450	porcine tra	352	40	46.5	334	2	Q3AMN9_SYNS9	Q3amn9	synecococc
280	41	47.7	1449	1	SPIKE_CVPMI	P33470	porcine tra	353	40	46.5	337	2	Q2V3S5_ARATH	Q2v3s5	arabidopsia
281	41	47.7	1449	2	Q4ZJU7_9CORO	Q4zju7	transmissib	354	40	46.5	337	2	Q7MYQ1_PHOLL	Q7myq1	photorhabd
282	41	47.7	1449	2	QBQW5_9CORO	QBqwu5	transmissib	355	40	46.5	338	2	Q5U496_XENLA	Q5u496	xenopus lae
283	41	47.7	1449	2	Q7T428_9CORO	Q7t428	transmissib	356	40	46.5	344	2	Q8A533_BACTN	Q8a533	bacteroides
284	41	47.7	1449	2	Q84853_9CORO	Q84853	porcine res	357	40	46.5	350	2	Q35LH7_9BRAD	Q35lh7	bradyrhizob
285	41	47.7	1449	2	Q88510_9CORO	Q88510	transmissib	358	40	46.5	358	2	Q638V1_BACC2	Q638v1	bacillus ce
286	41	47.7	1449	2	Q9D22_9CORO	Q9d22	transmissib	359	40	46.5	374	2	Q9N6B1_LEIMA	Q9n6b1	leishmania
287	41	47.7	1449	2	Q9YRA7_9CORO	Q9yra7	transmissib	360	40	46.5	375	2	Q4HS04_CAMUP	Q4hs04	campylobact
288	41	47.7	1449	2	Q2LJ83_9CORO	Q2lj83	transmissib	361	40	46.5	380	2	Q5ND66_ORILA	Q5nd66	oryzias lac
289	41	47.7	1451	1	SPIKE_CVCAI	P16300	canine ente	362	40	46.5	384	2	Q470K9_RALEU	Q470k9	raistonia e
290	41	47.7	1452	1	SPIKE_FIPV	P10033	feline infe	363	40	46.5	385	2	Q3RQY0_RALME	Q3rqy0	raistonia m
291	41	47.7	1452	2	Q4U5G0_9CORO	Q4u5g0	feline coro	364	40	46.5	385	2	Q3R7N9_RALEU	Q3r7n9	raistonia e
292	41	47.7	1452	2	Q52PA3_9CORO	Q52pa3	feline infe	365	40	46.5	385	2	Q5QYB3_IDILO	Q5qyb3	idimarina
293	41	47.7	1453	1	SPIKE_CVCAE	Q65984	canine ente	366	40	46.5	387	2	Q8XTI7_RALSO	Q8xti7	raistonia s
294	41	47.7	1453	2	Q6T522_9CORO	Q6t522	canine coro	367	40	46.5	388	2	Q7N1D5_PHOLL	Q7nid5	photorhabd
295	41	47.7	1453	2	Q6TPL2_9CORO	Q6tpl2	canine coro	368	40	46.5	389	2	Q82568_ARATH	Q82568	arabidopsia
296	41	47.7	1453	2	Q7T6T3_9CORO	Q7t6t3	canine coro	369	40	46.5	390	2	Q3JH09_BURP1	Q3jhq9	burkholderi
297	41	47.7	1453	2	Q5R264_9CORO	Q5r264	canine coro	370	40	46.5	390	2	Q2T373_BURTH	Q2t373	burkholderi
298	41	47.7	1454	2	Q66928_9CORO	Q66928	feline coro	371	40	46.5	390	2	Q62A68_BURMA	Q62a68	burkholderi
299	41	47.7	1454	2	Q6SR88_9CORO	Q6sr88	feline infe	372	40	46.5	390	2	Q63NU2_BURPS	Q63nu2	burkholderi
300	41	47.7	1472	2	Q87B49_XYLFT	Q87b49	xylella fas	373	40	46.5	393	2	Q46YU2_RALEU	Q46yu2	raistonia e
301	41	47.7	1521	2	Q3R2H6_XYLFA	Q3r2h6	xylella fas	374	40	46.5	398	2	Q221A8_RALEU	Q221a8	raistonia e
302	41	47.7	1521	2	Q3RA73_XYLFA	Q3ra73	xylella fas	375	40	46.5	400	2	Q21071_BORAV	Q21071	bordetella
303	41	47.7	1521	2	Q3RUW8_XYLFA	Q3ruw8	xylella fas	376	40	46.5	400	2	Q7VW96_BORPE	Q7vw96	bordetella
304	41	47.7	2157	2	Q9AYB5_ORYSA	Q9ayb5	oryza sativ	377	40	46.5	400	2	Q7W5Q6_BORPA	Q7w5q6	bordetella
305	41	47.7	2174	2	Q2QGT7_ORYSA	Q2qgt7	oryza sativ	378	40	46.5	400	2	Q7WD92_BORBR	Q7wd92	bordetella
306	41	47.7	4268	2	Q4QFM9_LEIMA	Q4qfm9	leishmania	379	40	46.5	410	2	Q3F1E2_BACTI	Q3fie2	bacillus th
307	41	47.7	6239	2	Q9S0R7_STRAW	Q9s0r7	streptomyc	380	40	46.5	422	2	Q5MIW4_AEDAL	Q5miw4	aedes albop
308	40.5	47.1	192	2	Q55940_9HIV1	O55940	human immun	381	40	46.5	435	2	Q4EUT3_LISMO	Q4eut3	listeria mo
309	40.5	47.1	192	2	Q55946_9HIV1	O55946	human immun	382	40	46.5	462	2	Q4H436_THETH	Q4h436	thermus the
310	40.5	47.1	192	2	Q55969_9HIV1	O55969	human immun	383	40	46.5	462	2	Q72HW2_THET2	Q72hw2	thermus the
311	40.5	47.1	250	1	Y2406_ANAP	Q44510	anabaena sp	384	40	46.5	462	2	Q6IMZ0_RAT	Q6imz0	rattus norv
312	40.5	47.1	271	2	Q3QH95_9GAMM	Q3qh95	shewanella	385	40	46.5	471	2	Q923M2_RAT	Q923m2	rattus norv
313	40.5	47.1	308	2	Q8FCZ9_ECOL6	Q8fcz9	escherichia	386	40	46.5	478	2	Q49QX8_SCHMA	Q49qx8	schistosoma
314	40.5	47.1	312	2	Q7XBE3_ORYSA	Q7xeb3	oryza sativ	387	40	46.5	485	2	Q5DHS0_SCHJA	Q5dhs0	schistosoma
315	40.5	47.1	319	2	Q2V079_CABEL	Q2v079	caenorhabdi	388	40	46.5	508	2	Q8N9M5_HUMAN	Q8n9m5	homo sapien
316	40.5	47.1	323	2	Q95X80_CABEL	Q95x80	caenorhabdi	389	40	46.5	529	2	Q61TI9_CAEBR	Q61ti9	caenorhabdi
317	40.5	47.1	324	2	O17172_CABEL	O17172	caenorhabdi	390	40	46.5	534	2	Q93YH2_LYCER	Q93yh2	lycopersico
318	40.5	47.1	413	2	Q2YCH3_NITMU	Q2ych3	nitrospir	391	40	46.5	541	2	Q3HTK2_CHLRE	Q3htk2	chlamydomon
319	40.5	47.1	74	2	Q9K0F7_NEIMB	Q9k0f7	neisseria m	392	40	46.5	563	2	Q4SCB8_TETNG	Q4scb8	tetradodon n
320	40	46.5	81	2	Q5FA07_NEIGI	O5fa07	neisseria g	393	40	46.5	572	2	Q8Y842_LISMO	Q8y842	listeria mo
321	40	46.5	81	2	Q9JVF9_NEINWA	Q9jvf9	neisseria m	394	40	46.5	578	2	Q47BW9_DSCAR	Q47bw9	dechloromon
322	40	46.5	105	2	Q3U595_MOUSE	Q3u595	mus musculus	395	40	46.5	592	2	Q7SY37_BRARE	Q7sy37	brachydanio
323	40	46.5	107	2	Q3WAP6_9ACTO	Q3wap6	frankia sp.	396	40	46.5					

397	40	46.5	626	2	Q5BCZ0_EMENI	Q5bcz0 aspergillus	470	39	45.3	240	2	Q9GTN2_DROSI	Q9gtn2 drosophila
398	40	46.5	634	2	Q2WIM4_CLOBE	Q2wim4 clostridium	471	39	45.3	240	2	Q9V402_DROME	Q9v402 drosophila
399	40	46.5	637	2	Q65WU5_ORYSA	Q65wu5 oryza sativ	472	39	45.3	241	2	Q3FSY8_RHOFER	Q3fsy8 rhodospira
400	40	46.5	658	2	Q88UJ1_LACPL	Q88uj1 lactobacilli	473	39	45.3	256	2	Q5ZNV2_9VIRU	Q5znv2 cotesia con
401	40	46.5	670	1	AREP_ARATH	Q93vr9 arabidopsis	474	39	45.3	261	1	CALB1_CHICK	P04354 gallus gall
402	40	46.5	677	2	Q7PP32_ANOGA	Q7pp32 anopheles g	475	39	45.3	262	2	Q8PX61_METWA	Q8px61 mechanosarc
403	40	46.5	686	2	Q3VAT3_9SPHN	Q3vat3 sphingopyxi	476	39	45.3	262	2	Q8TUK3_METAC	Q8tuk3 mechanosarc
404	40	46.5	696	2	Q2SG98_9GAMM	Q2sg98 shigella che	477	39	45.3	263	2	Q5TJQ21_HUMAN	Q5tjq21 homo sapien
405	40	46.5	720	2	Q16270_CABEL	Q16270 caenorhabdi	478	39	45.3	263	2	Q38UP2_LACSS	Q38up2 lactobacill
406	40	46.5	723	2	Q3GUH0_9ACTO	Q3guh0 nocardioid	479	39	45.3	285	2	Q9AIT5_VIBCH	Q9ait5 vibrio chol
407	40	46.5	723	2	Q97GM4_CLOBAB	Q97gm4 clostridium	480	39	45.3	288	2	Q7ZVN0_BRARE	Q7zvn0 brachydanio
408	40	46.5	724	2	Q9AD25_STRCO	Q9ad25 streptomyce	481	39	45.3	292	2	Q81DC5_BACCR	Q81dc5 bacillus ce
409	40	46.5	841	2	Q316Q4_DESDG	Q316q4 desulfovibr	482	39	45.3	294	1	WRK70_ARATH	Q91y00 arabidopsis
410	40	46.5	889	2	Q6PYI2_OSTTA	Q6pyi2 ostreococcu	483	39	45.3	295	2	Q4WLN0_ASPEU	Q4wln0 aspergillus
411	40	46.5	971	1	QOYIA_CABEL	Q09281 caenorhabdi	484	39	45.3	299	2	Q4ZP35_PSEU2	Q4zrp35 pseudomonas
412	40	46.5	1146	1	YHC3_YEAST	P38742 saccharomyc	485	39	45.3	305	2	Q62ZG8_BACLD	Q62zgj8 bacillus li
413	40	46.5	1162	2	Q7PMJ6_ANOGA	P38742 anopheles g	486	39	45.3	310	2	Q50J74_BRARE	Q50j74 brachydanio
414	40	46.5	1189	2	Q5JJA2_PYRKO	Q5jjj2 pyrococcus	487	39	45.3	313	2	Q2NR31_SODGL	Q2nr31 sodalis glo
415	40	46.5	1472	2	Q9PG24_XYLFa	Q9pg24 xyella fas	488	39	45.3	319	2	Q5NX35_AZOSE	Q5nx35 azoarcus sp
416	40	46.5	1614	2	Q5TRC9_ANOGA	Q5trcg anopheles g	489	39	45.3	319	2	Q6SP29_BACLD	Q6sp29 bacillus li
417	40	46.5	1675	2	Q582S2_9TRYP	Q582s2 trypanosoma	490	39	45.3	320	2	Q9LRQ2_ARATH	Q9lrg2 arabidopsis
418	39.5	45.9	192	2	Q4JGH0_9HIV1	Q4jgh0 human immun	491	39	45.3	321	2	Q66DG1_YERPS	Q66dgl yersinia ps
419	39.5	45.9	192	2	Q4U541_9HIV1	Q4u541 human immun	492	39	45.3	321	2	Q8ZDH6_YERPE	Q8zdh6 yersinia pe
420	39.5	45.9	257	2	Q953X6_9BILA	Q953x6 terebratali	493	39	45.3	330	2	Q5TIN8_BRARE	Q5tin8 brachydanio
421	39.5	45.9	310	2	Q72MN6_LEPTC	Q72mn6 leptospira	494	39	45.3	331	2	Q6KB56_TETNG	Q6kb56 tetraodon n
422	39.5	45.9	324	2	Q8EZA3_LEPIN	Q8eza3 leptospira	495	39	45.3	337	2	Q20269_CABEL	Q20269 caenorhabdi
423	39.5	45.9	324	2	Q47B91_DECAR	Q47b91 dechloromon	496	39	45.3	339	2	Q7UXZ2_RHOBA	Q7uxz2 rhodospirell
424	39.5	45.9	411	2	Q8QOM3_METWA	Q8qom3 methanosarc	497	39	45.3	343	1	FUT2_HYLLA	Q9ttc7 hylobates l
425	39.5	45.9	611	2	Q5RIF1_BRARE	Q5rif1 brachydanio	498	39	45.3	346	2	Q3E7K9_ARATH	Q3e7k9 arabidopsis
426	39.5	45.9	611	2	Q6NUX4_BRARE	Q6nux4 brachydanio	499	39	45.3	348	2	Q4QAL8_LEITWA	Q4qal8 leishmania
427	39.5	45.9	612	2	Q5ZJY6_CHICK	Q5zjj6 gallus gall	500	39	45.3	358	2	Q8U2T3_PYRFU	Q8u2t3 pyrococcus
428	39.5	45.9	630	2	Q4T8V9_TETNG	Q4t8v9 tetraodon n	501	39	45.3	365	2	Q59565_PYRHO	Q59565 pyrococcus
429	39.5	45.9	697	2	Q6CQK9_KLUJIA	Q6cqk9 kluyveromyc	502	39	45.3	365	2	Q9V213_PYRAB	Q9v213 pyrococcus
430	39.5	45.9	1867	2	Q2U2A2_ASPOR	Q2u2a2 aspergillus	503	39	45.3	370	2	Q2LQX0_9DELT	Q2lqx0 syntrophus
431	39	45.3	59	2	Q349J1_RHOPA	Q349j1 rhodopsendo	504	39	45.3	371	2	Q29819_ARCFU	Q29819 archaeoglob
432	39	45.3	59	2	Q34BD4_RHOPA	Q34bd4 rhodopsendo	505	39	45.3	371	2	Q59Y10_HUMAN	Q59y10 homo sapien
433	39	45.3	59	2	Q36VP4_RHOPA	Q36vp4 rhodopsendo	506	39	45.3	373	2	Q8AB77_BACTIN	Q8ab77 bacteroides
434	39	45.3	59	2	Q36VP6_RHOPA	Q36vp6 rhodopsendo	507	39	45.3	377	2	Q4ZVV0_PSEU2	Q4zvv0 pseudomonas
435	39	45.3	59	2	Q37BH8_RHOPA	Q37bh8 rhodopsendo	508	39	45.3	389	2	Q04178_BRACM	Q04178 brassica ca
436	39	45.3	59	2	Q37E67_RHOPA	Q37e67 rhodopsendo	509	39	45.3	389	2	Q82567_ARATH	Q82567 arabidopsis
437	39	45.3	59	2	Q37FH2_RHOPA	Q37fh2 rhodopsendo	510	39	45.3	389	2	Q39810_SOYBN	Q39810 glycine max
438	39	45.3	59	2	Q21ST5_RHOPA	Q21st5 rhodopsendo	511	39	45.3	389	2	Q3SP57_BRACM	Q3sp57 brassica ca
439	39	45.3	59	2	Q21W45_RHOPA	Q21w45 rhodopsendo	512	39	45.3	392	2	Q8WIE9_BRARP	Q8wie9 brassica ra
440	39	45.3	64	2	Q3EVA0_BACTI	Q3eva0 bacillus th	513	39	45.3	392	2	Q81A41_BACCR	Q81a41 bacillus ce
441	39	45.3	90	2	Q4R3Y9_TETNG	Q4r3y9 tetraodon n	514	39	45.3	400	2	Q7V7D6_PROMM	Q7v7d6 prochloroco
442	39	45.3	99	2	Q74LF6_LACJO	Q74lf6 lactobacill	515	39	45.3	402	2	Q4WCA8_ASPFU	Q4wca8 aspergillus
443	39	45.3	106	2	Q2W8L0_MAGSA	Q2w8l0 magnetospir	516	39	45.3	408	2	Q5BDP8_EMENI	Q5bdp8 aspergillus
444	39	45.3	108	2	Q9KEC8_BACHD	Q9kec8 bacillus ha	517	39	45.3	410	2	Q7CMD2_BACAN	Q7cmd2 bacillus an
445	39	45.3	133	2	Q2Z903_9GAMM	Q2z903 shewanella	518	39	45.3	410	2	Q9X367_BACILUS	Q9x367 bacillus an
446	39	45.3	135	2	Q05499_BACSU	Q05499 bacillus su	519	39	45.3	416	2	Q74NL9_BACCL	Q74nl9 bacillus bo
447	39	45.3	147	2	Q17986_CABEL	Q17986 caenorhabdi	520	39	45.3	417	2	Q31V01_SHIBS	Q31v01 shigella bo
448	39	45.3	159	2	Q7R5F3_GIALA	Q7r5f3 giardia lam	521	39	45.3	417	2	Q3YVY6_SHISS	Q3yvvy6 shigella so
449	39	45.3	161	2	Q3U6U3_MOUSE	Q3u6u3 mus musculu	522	39	45.3	417	2	Q8KMW4_ECOLI	Q8kmw4 escherichia
450	39	45.3	170	2	Q9VX72_DROME	Q9vx72 drosophila	523	39	45.3	417	2	Q9ZIT0_ECOLI	Q9zit0 escherichia
451	39	45.3	172	2	Q4MGF7_BACCE	Q4mgf7 bacillus ce	524	39	45.3	417	2	Q8FC97_ECOL6	Q8fc97 escherichia
452	39	45.3	173	2	Q8U2X6_PYRFU	Q8u2x6 pyrococcus	525	39	45.3	421	2	Q7UUD0_RHOPIRE	Q7uud0 rhodospirell
453	39	45.3	175	2	Q9VD84_DROME	Q9vd84 drosophila	526	39	45.3	422	2	Q5Z513_ORYSA	Q5z513 oryza sativ
454	39	45.3	179	2	Q5UQ20_HUMAN	Q5jq20 homo sapien	527	39	45.3	426	2	Q8A6V4_BACTN	Q8a6v4 bacteroides
455	39	45.3	183	2	Q55AD9_DICDI	Q55ad9 dictyosteli	528	39	45.3	427	2	Q5TQ40_ANOGA	Q5tq40 anopheles g
456	39	45.3	186	1	ARL6_MOUSE	Q88848 mus musculu	529	39	45.3	433	2	Q2SF93_9GAMM	Q2sf93 hanelia che
457	39	45.3	186	2	Q73QC5_TREDE	Q73qc5 treponema d	530	39	45.3	451	2	Q9LRQ7_ARATH	Q9lrg7 arabidopsis
458	39	45.3	186	2	Q3TUM2_MOUSE	Q3tum2 mus musculu	531	39	45.3	452	2	Q9FND9_ARATH	Q9fnp9 arabidopsis
459	39	45.3	186	2	Q8VCV3_MOUSE	Q8vcv3 mus musculu	532	39	45.3	455	2	Q5Z514_ORYSA	Q5z514 oryza sativ
460	39	45.3	187	2	Q684F5_9VIRU	Q684f5 sulfolobus	533	39	45.3	455	2	Q93XP5_ORYSA	Q93xp5 oryza sativ
461	39	45.3	193	2	Q3TV77_MOUSE	Q3ty77 mus musculu	534	39	45.3	458	2	Q82XT9_NITEU	Q8xt9 nitrosomona
462	39	45.3	197	2	Q5QSC8_9HIV1	Q5qsc8 human immun	535	39	45.3	471	2	Q3GYW7_9ACTO	Q3gyw7 nocardioid
463	39	45.3	199	2	Q84D96_9BACT	Q84d96 uncultured	536	39	45.3	477	2	Q5KQZ4_KLEPN	Q5kqz4 klebsiella
464	39	45.3	209	2	Q2L039_BORAV	Q2l039 bordetella	537	39	45.3	477	2	Q5MQC2_ECOLI	Q5mqc2 escherichia
465	39	45.3	213	2	Q2UTX6_ASPOR	Q2utx6 aspergillus	538	39	45.3	483	2	Q8XQP1_RALSO	Q8xqp1 raietonia s
466	39	45.3	218	2	Q518X8_PLAFA	Q5i8x8 plasmodium	539	39	45.3	493	2	Q43D94_CHLBR	Q43d94 chlorobium
467	39	45.3	218	2	Q8IINI_PLAFA	Q8iini plasmodium	540	39	45.3	504	2	Q5XL56_KLEPN	Q5xll56 klebsiella
468	39	45.3	230	2	Q6UCP8_9PROT	Q6ucp8 uncultured	541	39	45.3	515	2	Q9SIU7_ARATH	Q9siu7 arabidopsis
469	39	45.3	236	2	Q4PB40_METWA	Q4pe40 ustilago ma	542	39	45.3	523	2	Q8Z5B6_STRAW	Q8z5b6 streptomyce

543	39	45.3	534	1	TYRPL1_AMBME	P55027 ambystoma m	616	39	45.3	1003	2	Q6VLI9_HUMAN	Q6vli9 homo sapien
544	39	45.3	547	2	Q6PBV6_ACIAD	Q6fbv6 acinetobact	617	39	45.3	1035	2	Q69ZV9_MOUSE	Q69zv9 mus musculu
545	39	45.3	551	1	ASLA_ECOLI	P25549 escherichia	618	39	45.3	1050	2	Q50XRS_ENTHI	Q50xrs entamoeba h
546	39	45.3	551	2	Q329X6_SHIDS	Q229x6 shigella dy	619	39	45.3	1072	2	Q4SL50_TETNG	Q4sl50 tetraodon n
547	39	45.3	551	2	Q2M8A6_ECOLI	Q2m8a6 escherichia	620	39	45.3	1085	2	Q68D12_HUMAN	Q68d12 homo sapien
548	39	45.3	551	2	Q8XAQ0_ECO57	Q8xaq0 escherichia	621	39	45.3	1137	1	KCMAL1_CHICK	Q8ay88 g calcium-a
549	39	45.3	554	2	Q72MY0_LEPIC	Q72my0 leptospira	622	39	45.3	1144	2	Q9W7J2_TRASC	Q9w7j2 trachemys s
550	39	45.3	554	2	Q8EZI3_LEPIN	Q8ezi3 leptospira	623	39	45.3	1151	1	KCMAL1_MACMU	Q18867 m calcium-a
551	39	45.3	559	2	Q500V2_ARATH	Q500v2 arabidopsis	624	39	45.3	1152	1	KCMAL1_PIG	Q18866 s calcium-a
552	39	45.3	562	2	Q7QCQ5_ANOHA	Q7qcq5 anopheles g	625	39	45.3	1165	2	Q6D1L4_ERWCT	Q6d1l4 erwinia car
553	39	45.3	583	2	Q9VHN8_DROME	Q9vhn8 drosophila	626	39	45.3	1166	1	KCMAL1_BOVIN	Q28204 b calcium-a
554	39	45.3	590	2	Q59FH2_HUMAN	Q59fh2 homo sapien	627	39	45.3	1166	2	Q5SQR7_HUMAN	Q5sq77 homo sapien
555	39	45.3	591	1	Y1280_MYCTU	P66771 mycobacteri	628	39	45.3	1166	2	Q73728_TRASC	Q73728 trachemys s
556	39	45.3	591	1	Y1311_MYCBO	P66772 mycobacteri	629	39	45.3	1167	2	Q5SQ81_HUMAN	Q5sq81 homo sapien
557	39	45.3	595	2	Q59VY9_CANAL	Q59vy9 candida alb	630	39	45.3	1171	2	Q5SQ8C_HUMAN	Q5sq8c homo sapien
558	39	45.3	605	2	Q2ZXZ1_STRSU	Q2zxz1 streptococc	631	39	45.3	1173	2	Q73729_TRASC	Q73729 trachemys s
559	39	45.3	607	2	Q4ZLQ6_PSEU2	Q4zlg6 pseudomonas	632	39	45.3	1177	2	Q5SQ8R_HUMAN	Q5sq8r homo sapien
560	39	45.3	618	2	Q9HIW0_THEAC	Q9hiw0 thermoplasm	633	39	45.3	1178	2	Q5JQ23_HUMAN	Q5jq23 homo sapien
561	39	45.3	621	2	Q4QH29_LEIMA	Q4qhz9 leishmania	634	39	45.3	1179	1	KCMAL1_RABIT	Q9bg98 o calcium-a
562	39	45.3	624	2	Q65VV9_NANSM	Q65vv9 manheimia	635	39	45.3	1181	2	Q5SQ8B_HUMAN	Q5sq8b homo sapien
563	39	45.3	626	2	Q9VKU1_DROME	Q9vku1 drosophila	636	39	45.3	1181	2	Q9KN45_VIBCH	Q9kn45 vibrio chol
564	39	45.3	627	2	Q3AFG3_CARHZ	Q3afg3 carboxydoth	637	39	45.3	1192	2	Q3IT35_NATPD	Q3it35 natronomona
565	39	45.3	637	2	Q8BIG5_MOUSE	Q8big5 mus musculu	638	39	45.3	1196	1	KCMAL1_XENLA	Q90zc7 x calcium-a
566	39	45.3	652	2	Q4NLW4_WMICC	Q4nlw4 arthrobacte	639	39	45.3	1196	2	Q7NOR1_CHRVO	Q7nqr1 chromobacte
567	39	45.3	656	2	Q8CLP5_YERPE	Q8clp5 yersinia pe	640	39	45.3	1200	2	Q73731_TRASC	Q73731 trachemys s
568	39	45.3	662	2	Q6MU55_MYCMS	Q6mu55 mycoplasma	641	39	45.3	1203	1	PHYC2_HUMAN	Q9y6c5 homo sapien
569	39	45.3	669	2	Q8KKW5_RHIET	Q8kkw5 rhizobium e	642	39	45.3	1203	2	Q53Z57_HUMAN	Q53z57 homo sapien
570	39	45.3	672	2	Q66F41_YERPS	Q66f41 yersinia ps	643	39	45.3	1205	2	Q5SQ8S_HUMAN	Q5sq8s homo sapien
571	39	45.3	672	2	Q8ZBE2_YERPE	Q8zbe2 yersinia pe	644	39	45.3	1209	1	KCMAL1_MOUSE	Q08460 m calcium-a
572	39	45.3	686	2	Q3G6I3_DDELT	Q3g6i3 pelobacter	645	39	45.3	1209	2	Q5SQ8Z_HUMAN	Q5sq8z homo sapien
573	39	45.3	696	2	Q7S4L8_NEUCR	Q7s4l8 neuropsora	646	39	45.3	1210	1	KCMAL1_RAT	Q62976 r calcium-a
574	39	45.3	699	2	Q43TV5_SOLUS	Q43tv5 solibactera	647	39	45.3	1210	2	Q5SQ8A_HUMAN	Q5sq8a homo sapien
575	39	45.3	745	2	Q8K1O3_MOUSE	Q8k1o3 mus musculu	648	39	45.3	1222	2	Q7MFT6_VIBVY	Q7mtf6 vibrio vuln
576	39	45.3	752	2	Q3I1S8_PSEHT	Q3i1s8 pseudocalter	649	39	45.3	1222	2	Q8D4B1_VIBVU	Q8d4b1 vibrio vuln
577	39	45.3	753	2	Q3EHP2_ACTSC	Q3ehf2 actinobacil	650	39	45.3	1225	2	Q5SQ8Q_HUMAN	Q5sq8q homo sapien
578	39	45.3	756	2	Q2SPI3_GGAMM	Q2spi3 habella che	651	39	45.3	1231	2	Q73730_TRASC	Q73730 trachemys s
579	39	45.3	762	2	Q30368_RALEU	Q30368 ralstonia e	652	39	45.3	1236	1	KCMAL1_HUMAN	Q12791 h calcium-a
580	39	45.3	762	2	Q7WX97_RALEU	Q7wx97 ralstonia e	653	39	45.3	1249	2	QSRFV1_BRARE	Q5rfv1 brachydanio
581	39	45.3	773	2	Q3UQ23_MOUSE	Q3uq23 mus musculu	654	39	45.3	1251	2	Q4S3G4_TETNG	Q4s3g4 tetraodon n
582	39	45.3	807	2	Q6CRP5_YARLI	Q6crf5 yarrowia li	655	39	45.3	1289	2	Q17174_BOOMI	Q17174 boophilus m
583	39	45.3	809	1	IL4RA_HORSE	Q6wg24 equus caball	656	39	45.3	1320	2	Q7XXG1_ORYSA	Q7xxg1 oryza sativ
584	39	45.3	826	2	Q2ZMU9_SHEPU	Q2zmu9 shewanella	657	39	45.3	1337	2	Q53KQ8_ORYSA	Q53kq8 oryza sativ
585	39	45.3	826	2	Q366Z1_GGAMM	Q366z1 shewanella	658	39	45.3	1428	2	Q7Q867_ANOGA	Q7q867 anopheles g
586	39	45.3	833	2	Q9U2S8_CAEEL	Q9u2s8 caenorhabdi	659	39	45.3	1511	2	Q56A10_MOUSE	Q56a10 mus musculu
587	39	45.3	837	2	Q9HTP2_PSEAE	Q9htp2 pseudomonas	660	39	45.3	1534	2	Q9ULD9_HUMAN	Q9uld9 homo sapien
588	39	45.3	880	2	Q6PLP7_CHLRE	Q6plp7 chlamydomon	661	39	45.3	1756	2	Q61TS2_CAEBR	Q61ts2 caenorhabdi
589	39	45.3	885	2	Q5JQ19_HUMAN	Q5jq19 homo sapien	662	39	45.3	1801	2	Q2M085_DROPS	Q2m085 drosophila
590	39	45.3	889	2	Q83T68_SALTI	Q83t68 salmonella	663	39	45.3	1919	2	Q29S18_RABIT	Q29s18 oryctolagus
591	39	45.3	889	2	Q9X6X6_SALMON	Q9x6x6 salmonella	664	39	45.3	1920	2	Q29S19_RABIT	Q29s19 oryctolagus
592	39	45.3	889	2	Q57M84_SALCH	Q57m84 salmonella	665	39	45.3	1926	1	LPH_RABIT	P09848 oryctolagus
593	39	45.3	889	2	Q5PCP1_SALPA	Q5pcp1 salmonella	666	39	45.3	1927	1	LPH_HUMAN	P09848 homo sapien
594	39	45.3	889	2	Q8Z563_SALTI	Q8z563 salmonella	667	39	45.3	1927	2	Q4ZG58_HUMAN	Q4zg58 homo sapien
595	39	45.3	889	2	Q8ZNH2_SALTY	Q8znh2 salmonella	668	39	45.3	3213	2	Q4HZN3_GIBZE	Q4hzn3 gibberella
596	39	45.3	890	1	YOJN_ECOLI	P39838 escherichia	669	38.5	44.8	192	2	O55934_9HIV1	O55934 human immun
597	39	45.3	890	2	Q31239_SHIBS	Q31239 shigella bo	670	38.5	44.8	295	2	Q3QVU5_9RHOB	Q3qvus silicibacte
598	39	45.3	890	2	Q32119_SHIGL	Q32119 shigella dy	671	38.5	44.8	504	2	Q2TBU6_BOVIN	Q2tbus bos taurus
599	39	45.3	890	2	Q3Y2Z0_SHISS	Q3y2z0 shigella so	672	38.5	44.8	559	2	Q4EFU4_LISMO	Q4efu4 listeria mo
600	39	45.3	890	2	Q7UC77_SHIFL	Q7uc77 shigella fl	673	38.5	44.8	563	2	Q4ESB5_LISMO	Q4esb5 listeria mo
601	39	45.3	890	2	Q83Q06_SHIFL	Q83q06 shigella fl	674	38.5	44.8	563	2	Q71ZV4_LISMF	Q71zv4 listeria mo
602	39	45.3	890	2	Q8FPQ0_ECOL6	Q8fpq0 escherichia	675	38.5	44.8	563	2	Q2BY77_LISMO	Q2by77 listeria in
603	39	45.3	890	2	Q8XE40_ECO57	Q8xe40 escherichia	676	38.5	44.8	563	2	Q8Y7B8_LISMO	Q8y7b8 listeria mo
604	39	45.3	910	2	Q3XLH5_PPROT	Q3xlh5 magnetococc	677	38.5	44.8	609	1	LKHA4_RAT	P30349 rattus norv
605	39	45.3	913	2	Q4N589_THERA	Q4n589 theileria p	678	38.5	44.8	610	1	LKHA4_CAVPO	P19602 cavia porce
606	39	45.3	915	2	Q75G97_ORYSA	Q75g97 oryza sativ	679	38.5	44.8	610	1	LKHA4_CHILA	Q69sc8 chinchilla
607	39	45.3	924	2	Q2QRT7_ORYSA	Q2qrt7 oryza sativ	680	38.5	44.8	610	1	LKHA4_HUMAN	P09960 homo sapien
608	39	45.3	925	2	O5ARA5_EMENI	O5ara5 aspergillus	681	38.5	44.8	611	2	LKHA4_MOUSE	P24527 mus musculu
609	39	45.3	952	2	Q6Q0N0_RAT	Q6q0n0 rattus norv	682	38.5	44.8	611	2	Q6IAT6_HUMAN	Q6iat6 homo sapien
610	39	45.3	959	2	Q7TS67_MOUSE	Q7ts67 mus musculu	683	38.5	44.8	611	2	Q5REQ3_PONPY	Q5req3 pongo pygna
611	39	45.3	976	2	Q6FKX5_CANGA	Q6fkx5 candida gla	684	38.5	44.8	611	2	Q3SZHT_BOVIN	Q3szht bos taurus
612	39	45.3	979	1	CSTNI_MOUSE	Q9ep12 mus musculu	685	38.5	44.8	611	2	Q3UY71_MOUSE	Q3uy71 mus musculu
613	39	45.3	979	2	Q3HGG9_TRIER	Q3hgg9 trichodesmi	686	38.5	44.8	611	2	Q499P2_RAT	Q499p2 rattus norv
614	39	45.3	994	2	O5SLY5_THET8	O5sly5 thermus the	687	38	44.2	36	2	Q4XFV7_PLACHD	Q4xfv7 plasmodium
615	39	45.3	994	2	Q72GN0_THET2	Q72gm0 thermus the	688	38	44.2	48	2	Q855D7_9CAUD	Q855d7 mycobacteri

689	38	44.2	49	2	Q855N4_9CAUD	Q855N4 mycobacteri	762	38	44.2	289	1	AT15L5L_ARATH	Q9fir0 arabidopei
690	38	44.2	68	2	Q6K2B5_ORISA	Q6k2b5 oryza sativ	763	38	44.2	298	2	Q7XAT5_VICFA	Q7xat5 vicia faba
691	38	44.2	78	2	Q3EVG7_BACTI	Q3evg7 bacillus th	764	38	44.2	301	2	Q4K982_PSEF5	Q4k982 pseudomonas
692	38	44.2	88	2	Q56F07_9CAUD	Q56f07 aeromonas p	765	38	44.2	301	2	Q6MLG1_BDEBA	Q6mlg1 bdellovibri
693	38	44.2	88	2	Q6U9U8_9CAUD	Q6u9u8 aeromonas p	766	38	44.2	308	2	Q49X95_STAS1	Q49x95 staphylococ
694	38	44.2	89	2	Q8GC20_LEUME	Q8gc20 leuconostoc	767	38	44.2	309	2	Q54G71_DICDI	Q54g71 dictyosteli
695	38	44.2	105	2	Q6PSX5_HUMAN	Q6psx5 homo sapien	768	38	44.2	310	2	Q8XT83_RALSO	Q8xt83 raistonia s
696	38	44.2	105	2	Q5RE30_PONPY	Q5re30 pongo pygma	769	38	44.2	314	2	Q54DH5_DICDI	Q54dh5 dictyosteli
697	38	44.2	107	2	Q5Z5F0_ORYSA	Q5z5f0 oryza sativ	770	38	44.2	317	2	Q93YE3_TOBAC	Q93ye3 nicotiana t
698	38	44.2	110	2	Q31Q08_SYNP7	Q31q08 synechococc	771	38	44.2	319	2	Q2JC25_9ACTO	Q2jc25 frankia sp.
699	38	44.2	114	2	Q8Q6E9_9HIV1	Q8q6e9 human immun	772	38	44.2	324	2	Q7NPK5_GLOVI	Q7npk5 gloeobacter
700	38	44.2	115	2	Q7ZCZ5_9HIV1	Q7zc25 human immun	773	38	44.2	325	2	Q68334_VIBCH	Q68334 vibrio chol
701	38	44.2	121	2	Q8CC75_MOUSE	Q8cc75 mus musculu	774	38	44.2	325	2	Q8D4Y4_VIBVU	Q8d4y4 vibrio vuln
702	38	44.2	136	2	Q7UQE9_RHOBA	Q7uqe9 rhodopirell	775	38	44.2	326	2	Q3QER6_9GAMM	Q3qer6 shewanella
703	38	44.2	136	2	Q9X8X1_STRCO	Q9x8x1 streptomyce	776	38	44.2	327	2	Q9KTS1_VIBCH	Q9kts1 vibrio chol
704	38	44.2	137	2	Q4CJ34_CLOTRM	Q4cj34 clostridium	777	38	44.2	334	2	Q5ND67_ORYLA	Q5nd67 oryzias lat
705	38	44.2	144	2	Q2ZD10_9GAMW	Q2zdi0 shewanella	778	38	44.2	335	2	Q2VMQ7_ASPPU	Q2vmq7 aspergillus
706	38	44.2	144	2	Q8RGG9_FUSNN	Q8rgg9 fusobacteri	779	38	44.2	335	2	Q4WIX5_ASPPU	Q4wix5 aspergillus
707	38	44.2	150	2	Q5LLM9_SILPO	Q5llm9 silicibacte	780	38	44.2	340	2	Q5OJ70_BRARE	Q5oj70 brachydanio
708	38	44.2	162	2	Q7F999_ORYSA	Q7f999 oryza sativ	781	38	44.2	340	2	Q561V1_BRARE	Q561v1 brachydanio
709	38	44.2	162	2	Q7XR88_ORYSA	Q7xr88 oryza sativ	782	38	44.2	341	2	Q6EV32_BRARE	Q6ev32 brachydanio
710	38	44.2	164	2	Q7P4Z6_FUSNV	Q7p4z6 fusobacteri	783	38	44.2	342	2	Q33HD4_METHU	Q33hd4 methanospir
711	38	44.2	169	2	Q3C4C0_9CLOT	Q3c4c0 alkaliphilu	784	38	44.2	342	2	Q6NN16_DROME	Q6nn16 drosophila
712	38	44.2	170	2	Q2TIF9_ASPOP	Q2tyf9 aspergillus	785	38	44.2	342	2	Q8INF2_DROME	Q8inf2 drosophila
713	38	44.2	172	2	Q7QG36_ANOGA	Q7qg36 anopheles g	786	38	44.2	343	2	Q2UQT9_ASPOP	Q2uqt9 aspergillus
714	38	44.2	181	2	Q4N5P5_THEPA	Q4n5p5 theileria p	787	38	44.2	344	2	Q32YJ9_9EUKA	Q32yj9 ammonia sp.
715	38	44.2	181	2	Q4UDU9_THEAN	Q4udu9 theileria a	788	38	44.2	345	2	Q51932_PEPMA	Q51932 peptostrept
716	38	44.2	182	2	Q5CIZ1_CRYHO	Q5ciz1 cryptospori	789	38	44.2	350	2	Q5OJ63_FUGRU	Q5oj63 fugu rubrip
717	38	44.2	182	2	Q8DSR4_STRMU	Q8dsr4 streptococc	790	38	44.2	350	2	Q5TIN7_TETNG	Q5tin7 tetraodon n
718	38	44.2	183	2	Q7QVN8_GIALA	Q7qvn8 giardia lam	791	38	44.2	351	2	Q7O2S1_FUGRU	Q7o2s1 fugu rubrip
719	38	44.2	185	2	Q9VE56_DROME	Q9ve56 drosophila	792	38	44.2	352	1	DHSO_BACSU	Q60004 bacillus su
720	38	44.2	186	1	ARL6_HUMAN	Q9h0f7 homo sapien	793	38	44.2	356	2	Q82B48_STRAW	Q82b48 streptomyce
721	38	44.2	186	1	ARL6_PONPY	Q5r4g5 pongo pygma	794	38	44.2	357	2	Q31E22_THICR	Q31e22 thiomicrosp
722	38	44.2	186	2	Q5M9P8_BRARE	Q5m9p8 brachydanio	795	38	44.2	358	2	Q4KMS4_HUMAN	Q4kms4 homo sapien
723	38	44.2	187	2	Q4CCK1_CLOTRM	Q4cck1 clostridium	796	38	44.2	361	2	Q5U3S4_BRARE	Q5u3s4 brachydanio
724	38	44.2	188	2	Q49WF9_STAS1	Q49wf9 staphylococ	797	38	44.2	363	2	Q3FYD7_9DELT	Q3fyd7 pelobacter
725	38	44.2	188	2	Q4S1O6_TETNG	Q4si106 tetraodon n	798	38	44.2	365	2	Q9KXK4_STRCO	Q9kxk4 streptomyce
726	38	44.2	193	2	Q68EX8_XENLA	Q68ex8 xenopus lae	799	38	44.2	370	2	Q58M48_9CAUD	Q58m48 cyanophage
727	38	44.2	194	2	Q6CC85_YARLI	Q6cc85 yarrowia li	800	38	44.2	370	2	Q4CJL3_CLOTRM	Q4cj13 clostridium
728	38	44.2	201	2	Q4A2V2_9PHYC	Q4a2v2 emiliania h	801	38	44.2	374	2	Q6EV31_BRARE	Q6ev31 brachydanio
729	38	44.2	209	2	Q3FEI3_9BURK	Q3fei3 burkholderi	802	38	44.2	374	2	Q5ND65_ORYLA	Q5nd65 oryzias lat
730	38	44.2	212	2	Q3RSK6_RALME	Q3rsk6 ralstonia m	803	38	44.2	375	1	ACT_GIALA	PS1775 giardia lam
731	38	44.2	212	2	Q88TQ8_LACPL	Q88tq8 lactobacill	804	38	44.2	375	2	Q2U7A3_ASPOP	Q2u7a3 aspergillus
732	38	44.2	214	2	Q3FRJ6_9BURK	Q3frj6 rhodofexax	805	38	44.2	375	2	Q7QVY2_GIALA	Q7qvY2 giardia lam
733	38	44.2	217	2	Q9KMV8_VIBCH	Q9kmv8 vibrio chol	806	38	44.2	380	2	Q8TOK7_DROME	Q8tOk7 drosophila
734	38	44.2	220	2	Q8SU59_ENCCU	Q8su59 encephalito	807	38	44.2	381	2	Q74D11_GEOSL	Q74d11 geobacter s
735	38	44.2	220	2	Q3ERY5_BACTI	Q3ery5 bacillus th	808	38	44.2	384	2	Q93YE7_TOBAC	Q93ye7 nicotiana t
736	38	44.2	224	2	Q53IY4_RHOS2	Q53iy4 rhodomonas	809	38	44.2	384	2	Q93YF0_TOBAC	Q93yf0 nicotiana t
737	38	44.2	224	2	Q9LKN8_GUITH	Q9lkn8 guillardia	810	38	44.2	385	1	YG1W_YEAST	PS3230 saccharomyc
738	38	44.2	225	2	Q9QDB1_9GEMI	Q9qdb1 cowpea gold	811	38	44.2	385	2	Q45U38_YEAST	Q45u38 saccharomyc
739	38	44.2	226	2	Q3VIB6_9CHLB	Q3vib6 pelodictyon	812	38	44.2	391	2	Q2UQR1_ASPOP	Q2uqr1 aspergillus
740	38	44.2	227	2	Q3YJ63_9LILI	Q3yj63 euterpe ole	813	38	44.2	396	2	Q82QT0_STRAW	Q82qt0 streptomyce
741	38	44.2	233	2	Q5OJ71_BRARE	Q5oj71 brachydanio	814	38	44.2	398	2	Q2ZLI0_SHEPU	Q2zli0 shewanella
742	38	44.2	246	2	Q499B8_BRARE	Q499b8 brachydanio	815	38	44.2	402	1	2322A_HUMAN	Q2zli0 shewanella
743	38	44.2	248	2	Q68SX1_9CIRC	Q68sx1 beak and fe	816	38	44.2	402	1	2322A_MACFA	Q4r7x8 macaca fasc
744	38	44.2	255	2	Q4V208_BACCZ	Q4v208 bacillus ce	817	38	44.2	402	2	Q4TIH3_TETNG	Q4tIh3 tetraodon n
745	38	44.2	256	2	Q3FH96_9BURK	Q3fhn6 burkholderi	818	38	44.2	404	2	Q871K2_NEUCR	Q871k2 neurospora
746	38	44.2	256	2	Q3J9721_BURS3	Q3j9721 burkholderi	819	38	44.2	410	1	2322A_MOUSE	Q8bz89 mus musculu
747	38	44.2	256	2	Q32234_BACSU	Q32234 bacillus su	820	38	44.2	410	2	Q3TDD5_MOUSE	Q3tdd5 mus musculu
748	38	44.2	259	2	Q34439_BACSU	Q34439 bacillus su	821	38	44.2	419	1	RBW4_BRARE	Q6iq97 brachydanio
749	38	44.2	267	1	YBX1_BACSU	P54427 bacillus su	822	38	44.2	420	2	Q8EL05_OCEIH	Q8el05 oceanobacil
750	38	44.2	269	2	Q53FD4_HUMAN	Q53fd4 homo sapien	823	38	44.2	423	2	Q49A28_HUMAN	Q49a28 homo sapien
751	38	44.2	269	2	Q2KSD7_ADE04	Q2ksd7 human adeno	824	38	44.2	425	2	Q93YB9_TOBAC	Q93yb9 nicotiana t
752	38	44.2	274	2	Q6H1B3_9ADEN	Q6h1b3 human adeno	825	38	44.2	425	2	Q93YD0_TOBAC	Q93yd0 nicotiana t
753	38	44.2	274	2	Q8BEL2_ADE04	Q8bel2 human adeno	826	38	44.2	425	2	Q93YD6_TOBAC	Q93yd6 nicotiana t
754	38	44.2	276	2	Q2STU2_BURTH	Q2stu2 burkholderi	827	38	44.2	425	2	Q9XG72_TOBAC	Q9xg72 nicotiana t
755	38	44.2	276	2	Q3JV81_BURPL	Q3jv81 burkholderi	828	38	44.2	426	2	Q93YB0_TOBAC	Q93ye0 nicotiana t
756	38	44.2	276	2	Q62G24_BURMA	Q62g24 burkholderi	829	38	44.2	438	2	Q2J527_9ACTO	Q2j527 frankia sp.
757	38	44.2	276	2	Q63PT4_BURPS	Q63pt4 burkholderi	830	38	44.2	439	2	Q67K70_SYMTH	Q67k70 symbiobacte
758	38	44.2	283	2	Q583W4_9TRYP	Q583w4 trypanosoma	831	38	44.2	442	2	Q9K9L6_BACHD	Q9k9l6 bacillus ha
759	38	44.2	285	2	Q4KN40_HUMAN	Q4kn40 homo sapien	832	38	44.2	450	2	Q4S1E0_TETNG	Q4s1e0 tetraodon n
760	38	44.2	285	2	Q9H9F7_HUMAN	Q9h9f7 homo sapien	833	38	44.2	452	2	Q7NMK8_GLOVI	Q7nmk8 gloeobacter
761	38	44.2	288	2	Q9ZXA4_BPPHC	Q9zxa4 bacteriopho	834	38	44.2	453	2	Q37ZF4_SPHAR	Q37zf4 novosphingo

835	38	44.2	456	2	Q3EVW8_BACTI	Q3evw8 bacillus th	908	38	44.2	898	2	Q7MUA9_PORGI	Q7mu9 porphyromon
836	38	44.2	469	2	Q7PYE2_ANOGA	Q7pye2 anopheles g	909	38	44.2	901	2	Q9VI72_DROME	Q9vi72 drosophila
837	38	44.2	481	2	Q6WJ8_SCHMA	Q6wj8 schistosoma	910	38	44.2	904	2	Q9SDG8_ORYSA	Q9sdg8 oryza sativ
838	38	44.2	481	2	Q7NA15_WYCGA	Q7nai5 mycoplasma	911	38	44.2	906	2	Q4Z052_PLABE	Q4z052 plasmodium
839	38	44.2	483	2	Q4PHK1_USTMA	Q4phk1 utillago ma	912	38	44.2	909	2	Q6FLK9_CANGA	Q6flk9 candida gla
840	38	44.2	487	2	Q76399_CABEL	Q76399 caenorhabdi	913	38	44.2	933	2	Q4H9L2_9DEIO	Q4h9l2 deinococcus
841	38	44.2	488	2	Q8A510_BACTN	Q8a510 bacteroides	914	38	44.2	964	1	CTA2_BACCI	Q710873 bacillus ci
842	38	44.2	491	2	Q5LYN0_STRT1	Q5lyn0 streptococc	915	38	44.2	980	2	Q982K5_RHILO	Q982k5 rhizobium l
843	38	44.2	491	2	Q5M393_STRT2	Q5m393 streptococc	916	38	44.2	997	2	Q4UDD9_THEAN	Q4udd9 theileria a
844	38	44.2	499	2	Q2NQB1_SODGL	Q2nqb1 sodalis glo	917	38	44.2	1007	2	Q376P3_RHOPA	Q376p3 rhodopseudo
845	38	44.2	503	2	Q2U7B5_ASPOR	Q2u7b5 aspergillus	918	38	44.2	1011	2	Q2UCZ1_ASPOR	Q2ucz1 aspergillus
846	38	44.2	503	2	Q44T25_CHLLI	Q44t25 chlorobium	919	38	44.2	1016	2	Q44AJ5_SOLUS	Q44aj5 solibacter
847	38	44.2	508	2	Q88W37_LACPL	Q88w37 lactobacill	920	38	44.2	1047	2	Q81PB9_DROME	Q81pb9 drosophila
848	38	44.2	517	2	Q2N587_9SPHN	Q2n5e7 erythro bact	921	38	44.2	1051	2	Q9VKX1_DROME	Q9vkk1 drosophila
849	38	44.2	519	2	Q49AM3_HUMAN	Q49am3 homo sapien	922	38	44.2	1055	2	Q5OXI7_ENTHI	Q5oxi7 entamoeba h
850	38	44.2	520	2	Q7VIB9_HELHP	Q7vib9 helicobacte	923	38	44.2	1103	2	Q5QX17_ENTHI	Q5qxc17 entamoeba h
851	38	44.2	521	2	Q4WM02_ASPFU	Q4wm02 aspergillus	924	38	44.2	1139	1	MA2A2_HUMAN	Q3gbd2 syntrophomo
852	38	44.2	525	2	Q5LN01_SILPO	Q5ln01 silicibacte	925	38	44.2	1182	2	Q7UJC2_RHOBA	P49641 homo sapien
853	38	44.2	528	2	Q3FW33_9BURK	Q3fw33 rhodofera x	926	38	44.2	1187	2	Q6NSM8_BRARE	Q7ujc2 rhodopirell
854	38	44.2	529	2	Q6QN29_NPVSP	Q6qn29 spodoptera	927	38	44.2	1189	2	Q6PHV1_BRARE	Q6nsm8 brachydanio
855	38	44.2	531	2	Q74131_YARLI	Q74131 yarrowia li	928	38	44.2	1233	2	Q75H40_ORYSA	Q6phv1 brachydanio
856	38	44.2	534	2	Q97XY4_SULSO	Q97xy4 sulfolobus	929	38	44.2	1238	2	Q5W6J6_ORYSA	Q75h40 oryza sativ
857	38	44.2	538	1	CP392_DROME	P82713 drosophila	930	38	44.2	1305	2	Q4SPN5_TETNG	Q5w6j6 oryza sativ
858	38	44.2	541	2	Q7PR27_ANOGA	Q7pr27 anopheles g	931	38	44.2	1326	2	Q8AW94_FUGRU	Q4spn5 tetraodon n
859	38	44.2	554	2	Q4T606_TETNG	Q4t606 tetraodon n	932	38	44.2	1346	2	Q3HA57_TRIER	Q8aw94 fugu rubrip
860	38	44.2	566	2	P77943_SULAC	P77943 sulfolobus	933	38	44.2	1437	2	Q5SI71_CRYNE	Q3ha57 trichodesmi
861	38	44.2	566	2	Q4JA37_SULAC	Q4ja37 sulfolobus	934	38	44.2	1437	2	Q5K7Q5_CRYNE	Q5si71 cryptococcu
862	38	44.2	575	2	Q4RID9_TETNG	Q4rid9 tetraodon n	935	38	44.2	1470	2	Q3J8W4_NITOC	Q5k7q6 cryptococcu
863	38	44.2	591	2	Q8AAV1_BACTN	Q8aav1 bacteroides	936	38	44.2	1573	2	Q3KDT3_PSEPF	Q3j8w4 nitrosococc
864	38	44.2	604	2	Q53PG0_ORYSA	Q53pg0 oryza sativ	937	38	44.2	1604	2	Q7XQ14_ORYSA	Q3kdt3 pseudomonas
865	38	44.2	609	2	Q3MY84_9DELT	Q3my84 syntrophoba	938	38	44.2	1845	2	Q8OUA8_MOUSE	Q7xq14 oryza sativ
866	38	44.2	612	2	Q8MU94_MUSDO	Q8mu94 musca domes	939	38	44.2	1845	2	Q5QNG9_MOUSE	Q8oua8 mus musculu
867	38	44.2	629	2	Q7QFG0_ANOGA	Q7qfg0 anopheles g	940	38	44.2	1855	2	Q8QZFO_RAT	Q5qng9 mus musculu
868	38	44.2	631	2	Q8EJ70_SHEON	Q8ej70 shewanella	941	38	44.2	1860	2	Q8IZC6_HUMAN	Q8qzf0 rattus norv
869	38	44.2	633	2	Q75VY0_9DIPT	Q75vy0 culex trita	942	38	44.2	2075	2	Q4RYE5_TETNG	Q8izc6 homo sapien
870	38	44.2	634	2	Q7UPL5_RHOBA	Q7upl5 rhodopirell	943	38	44.2	2194	2	Q9GQR0_DROME	Q4rye5 tetraodon n
871	38	44.2	638	2	Q3Q0K3_9GAMM	Q3q0k3 shewanella	944	38	44.2	2348	2	Q36WB4_RHOPA	Q9gqr0 drosophila
872	38	44.2	641	2	Q2X2K3_9GAMM	Q2x2k3 shewanella	945	38	44.2	2772	2	Q9AV4V_DROME	Q36we4 rhodopseudo
873	38	44.2	644	2	Q2ZBT0_9GAMM	Q2zbt0 shewanella	946	38	44.2	2776	2	Q869A0_DROME	Q9av4v drosophila
874	38	44.2	644	2	Q363P1_9GAMM	Q363p1 shewanella	947	38	44.2	2894	2	Q7KRX2_DROME	Q869a0 drosophila
875	38	44.2	645	2	Q7RTL9_ANOGA	Q7rtl9 anopheles g	948	38	44.2	2898	2	Q868Z9_DROME	Q7kxr2 drosophila
876	38	44.2	647	2	Q3ARE9_CHLCH	Q3are9 chlorobium	949	37.5	43.6	147	1	RS12_METMP	Q6lxi4 methanococc
877	38	44.2	648	2	Q36HC5_9GAMM	Q36hcs shewanella	950	37.5	43.6	162	2	Q9S394_LACHI	Q9s394 lactobacill
878	38	44.2	649	1	ACES_DROME	P07140 drosophila	951	37.5	43.6	192	2	Q5S994_9HIV1	Q5s994 human immun
879	38	44.2	650	2	Q5TS70_ANOGA	Q5ts70 anopheles g	952	37.5	43.6	192	2	Q4JGK4_9HIV1	Q4jgk4 human immun
880	38	44.2	658	2	Q6CD14_YARLI	Q6cd14 yarrowia li	953	37.5	43.6	192	2	Q5EEI6_9HIV1	Q5eei6 human immun
881	38	44.2	664	1	ACES_ANOST	P56161 anopheles s	954	37.5	43.6	192	2	Q8UTS6_9HIV1	Q8uts6 human immun
882	38	44.2	673	2	Q5KIT3_BACDO	Q5ktt3 bactrocera	955	37.5	43.6	192	2	Q2VAK8_9HIV1	Q2vak8 human immun
883	38	44.2	673	2	Q5QTL8_BACDO	Q5qtl8 bactrocera	956	37.5	43.6	199	1	SODM2_HALVO	Q03301 halobacteri
884	38	44.2	673	2	Q8MVZ4_BACOL	Q8mvz4 bactrocera	957	37.5	43.6	200	1	SODM1_HALVO	Q03300 halobacteri
885	38	44.2	678	2	Q5BF27_9MENI	Q5bf27 aspergillus	958	37.5	43.6	269	2	Q5CZ70_HUMAN	Q5cz70 homo sapien
886	38	44.2	681	2	Q8ZV43_PYRAE	Q8z4v3 pyrobaculum	959	37.5	43.6	293	2	Q7NTE7_CHRVO	Q7nte7 chromobacte
887	38	44.2	691	2	Q8MXC5_MUSDO	Q8mxc5 musca domes	960	37.5	43.6	307	2	Q894D9_CLOTE	Q894d9 clostridium
888	38	44.2	691	2	Q2YHQ7_MUSDO	Q2yhq7 musca domes	961	37.5	43.6	310	2	Q6RXB1_HCMV	Q6rxb1 human cytom
889	38	44.2	692	2	Q7YMW9_MUSDO	Q7ymw9 musca domes	962	37.5	43.6	310	2	Q6SVZ2_HCMV	Q6svz2 human cytom
890	38	44.2	692	2	Q8MXC4_MUSDO	Q8mxc4 musca domes	963	37.5	43.6	329	2	Q337Y9_ORYSA	Q337y9 oryza sativ
891	38	44.2	692	2	Q8MXC6_MUSDO	Q8mxc6 musca domes	964	37.5	43.6	362	2	Q7XEE0_ORYSA	Q7xee0 oryza sativ
892	38	44.2	692	2	Q8MXC7_MUSDO	Q8mxc7 musca domes	965	37.5	43.6	385	2	Q645R2_TARGR	Q645r2 taricha gra
893	38	44.2	692	2	Q8MXC8_MUSDO	Q8mxc8 musca domes	966	37.5	43.6	424	2	Q6SEB4_9CAUD	Q6seb4 lactobacill
894	38	44.2	692	2	Q8MXC9_MUSDO	Q8mxc9 musca domes	967	37.5	43.6	424	2	Q6SPU5_LACJO	Q6spu5 lactobacill
895	38	44.2	692	2	Q9SP20_MUSDO	Q9sp20 musca domes	968	37.5	43.6	440	2	Q6X2M1_9BIVA	Q6x2m1 chlamys far
896	38	44.2	692	2	Q95WV7_MUSDO	Q95wv7 musca domes	969	37.5	43.6	466	2	Q6ZQH9_MOUSE	Q6zqh9 mus musculu
897	38	44.2	704	2	Q3BDV8_HAEIR	Q3bdv8 haematobia	970	37.5	43.6	478	2	Q8C2R4_MOUSE	Q8c2r4 mus musculu
898	38	44.2	708	2	P91954_LUCCU	P91954 lucilia cup	971	37.5	43.6	546	2	Q3TO62_MOUSE	Q3tq62 mus musculu
899	38	44.2	714	2	Q5T1U7_HUMAN	Q5tlu7 homo sapien	972	37.5	43.6	546	2	Q66HD9_RAT	Q66hd9 rattus norv
900	38	44.2	730	2	Q4WNP3_ASPFU	Q4wnp3 aspergillus	973	37.5	43.6	548	1	IDD_MOUSE	P98154 mus musculu
901	38	44.2	756	2	Q39PX6_GEOGM	Q39px6 geobacter m	974	37.5	43.6	549	2	Q6P5A9_MOUSE	Q6p5a9 m digeorge
902	38	44.2	773	2	Q93FB8_9RHOH	Q93fb8 azoarcus ev	975	37.5	43.6	550	1	IDD_HUMAN	P98153 homo sapien
903	38	44.2	774	2	Q9IAV5_BRARE	Q9iav5 brachydanio	976	37.5	43.6	550	2	Q8IWC8_HUMAN	Q8iwc8 homo sapien
904	38	44.2	796	2	Q6OAT9_WETCA	Q6oat9 methylococc	977	37.5	43.6	550	2	Q5R6P9_PONPY	Q5r6p9 pongo pygma
905	38	44.2	798	1	HASP_HUMAN	Q8tcf76 homo sapien	978	37.5	43.6	554	2	Q21P51_9DELT	Q21p51 anaeronyxob
906	38	44.2	812	2	Q9IAV4_BRARE	Q9iav4 brachydanio	979	37.5	43.6	565	2	Q3FP38_RHOMR	Q3fp38 rhodothermu
907	38	44.2	887	2	Q3WPP2_9RHIZ	Q3wpp2 mesorhizobi	980	37.5	43.6	580	2	Q8CB23_MOUSE	Q8cb23 mus musculu

881 37.5 43.6 588 2 Q3Z7M9_DEHE1
 982 37.5 43.6 1081 2 Q6F8U9_AC1AD
 983 37.5 43.6 1241 2 Q80S25_NDIAD
 984 37.5 43.6 1248 2 Q80S39_VIRU
 985 37.5 43.6 1266 2 Q80S39_VIRU
 986 37 43.0 33 2 Q4R1H4_PIG
 987 37 43.0 47 2 Q70U88_9AGAR
 988 37 43.0 58 2 Q18895_CANFA
 989 37 43.0 59 2 Q37CK6_RHOPA
 990 37 43.0 59 2 Q2IRN0_RHOPA
 991 37 43.0 76 2 Q5YQ00_NOCPA
 992 37 43.0 86 1 VFW_CMVVS
 993 37 43.0 97 2 Q4Z9Q0_9CAUD
 994 37 43.0 103 2 Q4XJH8_PLACH
 995 37 43.0 107 2 Q4R1H5_PIG
 996 37 43.0 108 2 Q5ALN4_CANAL
 997 37 43.0 109 2 Q6C043_YARLI
 998 37 43.0 110 2 Q38E29_9TRYP
 999 37 43.0 113 2 Q77162_ENTHI
 1000 37 43.0 113 2 Q77163_ENTIV

Q3z7m9 dehalococco
 Q6f8u9 acinetobact
 Q80s25 ndum virus
 Q80s39 bebaru viru
 Q80s39 bebaru viru
 Q9laa1 staphylococ
 Q4r1h4 sus scrofa
 Q70u88 hebeloma ra
 Q18895 canis famil
 Q37ck6 rhodopseudo
 Q2irn0 rhodopseudo
 Q5ywq0 nocardia fa
 P19510 ovine lent
 Q4z9q0 bacterioph
 Q4xjh8 plasmodium
 Q4r1h5 sus scrofa
 Q5aln4 candida alb
 Q6c043 yarrowia li
 Q38e29 trypanosoma
 Q77162 entamoeba h
 Q77163 entamoeba h

ALIGNMENTS

RESULT 1
 HPSE HUMAN
 ID HPSE HUMAN STANDARD; PRT; 543 AA.
 AC Q9Y5L1; Q5G6S; Q9UL39;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-NOV-1999, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
 DE subunit].
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRL, HPSE1, HSE1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
 RA Seddon A.P., Giorgio N.A., Bohlen P.;
 RT "Cloning and functional expression of a human heparanase gene.";
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.
 RC TISSUE=Embryonic fibroblast;
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
 RA Toyoshima M., Nakajima M.;
 RT "Human heparanase. Purification, characterization, cloning, and
 RT expression.";
 RL J. Biol. Chem. 274:24153-24160(1999).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
 RX PubMed=10395325; DOI=10.1038/10518;
 RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
 RA Spector L., Pecker I.;
 RT "Mammalian heparanase: gene cloning, expression and function in tumor
 RT progression and metastasis.";
 RL Nat. Med. 5:793-802(1999).
 RN [4]
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
 RP OF 158-174; 263-272; 326-337; 433-436; 466-468 AND 478-483.
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 RN [5]
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=2029546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage.";
 RL Glycobiology 10:467-475(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RX PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin I., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis.";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
 RN [7]
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Parekh R., Stuberfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpal) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435(2003).
 RN [8]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RA Pinhal M.A., Semedo P.;
 RT "Cloned heparanase from MCF-7 cells.";
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Small intestine;
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
 RA Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Sapletton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [11]
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RX PubMed=11123890; DOI=10.1021/bi002080p;
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Greedy J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic
 RT endoglycosidase heparanase".
 RL Biochemistry 39:15459-15667(2000).
 RN [12]
 RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
 RP ASN-217; ASN-238 AND ASN-459.
 RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
 RA Samizu S., Ishida K., Wierzb M.K., Osada H.;
 RA "Secretion of heparanase protein is regulated by glycosylation in
 RT human tumor cell lines.";
 RT J. Biol. Chem. 279:2697-2703(2004).
 RN [13]
 RN SUBCELLULAR LOCATION.
 RP PubMed=15292202; DOI=10.1074/jbc.M402131200;
 RX Gingis-Velietki S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
 RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
 RA "Heparanase uptake is mediated by cell membrane heparan sulfate
 RT proteoglycans.";
 RT J. Biol. Chem. 279:44084-44092(2004).
 RN [14]
 RP BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
 RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
 RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
 RA "Heparanase processing by lysosomal/endosomal protein preparation.";
 RL FEBS Lett. 579:2334-2338(2005).
 RN [15]
 RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
 RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
 RA Abboud-Jarrou G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
 RA Elgavish S., Peretz T., Vlodavsky I.;
 RA "Site-directed mutagenesis, proteolytic cleavage, and activation of
 RT human proheparanase.";
 RT J. Biol. Chem. 280:13568-13575(2005).
 RN [16]
 RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
 RP PubMed=15769092; DOI=10.1074/jbc.M414546200;
 RX Levy-Adam F., Abboud-Jarrou G., Guerrini M., Beccati D.,
 RA Vlodavsky I., Ilan N.;
 RA "Identification and characterization of heparin/heparan sulfate
 RT binding domains of the endoglycosidase heparanase.";
 RT J. Biol. Chem. 280:20457-20466(2005).
 RN [17]
 RP VARIANT SER-260.
 RX PubMed=15334672;
 RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
 RA Shi L.H., Shen H.W., Xu B.;
 RA "Heparanase mRNA expression and point mutation in hepatocellular
 RT carcinoma.";
 RT World J. Gastroenterol. 10:2795-2799(2004).
 RL
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Due to its contribution to metastasis and
 CC angiogenesis, it is considered to be a potential target for anti-
 CC cancer therapies.
 CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
 CC lower extent, by heparin and sulfamin and activated by calcium and
 CC magnesium (By similarity).
 CC -1- BIOPHYSICOCHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 4-6;
 CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products.
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium.
 CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
 CC weakly expressed in lymph node, thymus, peripheral blood
 CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.
 CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme.
 CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility.

Query Match 100.0%; Score 86; DB 1; Length 543;
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14
 |||||
 Db 294 TWHYYLNGRTATR 307

RESULT 2

ID HPSE_CHICK STANDARD; PRT: 523 AA.
 AC Q90YK5;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-DEC-2001, sequence version 1.
 DT 07-FEB-2006, entry version 12.
 DE Heparanase precursor (EC 3.2.-.-).
 GN Name=HPSE; Synonyms=HPA;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RX MEDLINE=2136959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
 RA Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
 RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
 RT "Expression pattern and secretion of human and chicken heparanase are
 RT determined by their signal peptide sequence.";
 RL J. Biol. Chem. 276:29178-29187(2001).
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans (By similarity).
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
 CC (By similarity).
 CC -1- PTM: N-glycosylated (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NonDerivs License

 CC EMBL: AY017007; AAK82648.1; -; mRNA.
 CC Ensembl: ENSGALG00000011203; Gallus gallus.
 CC InterPro: IPR005199; Glyco_hydro_79_N.
 CC Pfam: PF03662; Glyco_hydro_79n; 1.
 CC DR Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
 FT SIGNAL 1 18 Potential.
 FT CHAIN 19 523 Heparanase.
 FT FTID=PRO_0000042259.
 FT REGION 137 141 Heparin/HS-binding (By similarity).
 FT REGION 250 260 Heparin/HS-binding (By similarity).
 FT ACT_SITE 204 204 Proton donor (Potential).
 FT ACT_SITE 323 323 Nucleophile (Potential).
 FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF81 CRC64;

Query Match 91.9%; Score 79; DB 1; Length 523;
 Best Local Similarity 85.7%; Pred. No. 0.00064;
 Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 DB 274 TWHYYVNGRSATR 287

RESULT 3
 HPSE_RAT STANDARD; PRT; 536 AA.
 ID HPSE_RAT
 AC Q71RP1; Q90ZF8;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 05-JUL-2004, sequence version 1.
 DT 07-MAR-2006, entry version 11.
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
 GN Name:Hpse; Synonyms:Hep;
 OS Rattus norvegicus (Rat);
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
 RA Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
 RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;
 RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
 RA "Characterization of heparanase from a rat parathyroid cell line.";
 RL J. Biol. Chem. 277:32459-32465(2002).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Contributes to metastasis and angiogenesis
 CC (By similarity).
 CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
 CC extent, by heparin and sulfamin (By similarity). Activated by
 CC calcium and magnesium. Inhibited by EDTA.
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products (By similarity).
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium
 CC (By similarity).
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility (By similarity).
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL: AF359508; AAQ15189.1; -; mRNA.
 CC EMBL: AF184967; AAF04563.1; -; mRNA.
 CC RGD: 61969; Hpse.
 CC InterPro: IPR005199; Glyco_hydro_79_N.
 CC Pfam: PF03662; Glyco_hydro_79n; I.
 CC Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;

KW Signal.
 FT CHAIN 1 28 By similarity.
 FT CHAIN 29 102 Heparanase 8 kDa subunit.
 FT PROPEP 103 150 /FTid=PRO_0000042266.
 FT CHAIN 151 536 Linker peptide (By similarity).
 FT CHAIN 151 536 /FTid=PRO_0000042267.
 FT CHAIN 151 536 Heparanase 50 kDa subunit.
 FT REGION 151 155 /FTid=PRO_0000042268.
 FT REGION 263 273 Heparin/HS-binding (By similarity).
 FT ACT_SITE 218 218 Heparin/HS-binding (By similarity).
 FT ACT_SITE 336 336 Proton donor (Potential).
 FT CARBOHYD 155 155 Nucleophile (Potential).
 FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).
 FT CONFLICT 15 15 G -> R (in Ref. 2).
 FT CONFLICT 227 227 H -> Q (in Ref. 2).
 FT CONFLICT 350 350 D -> N (in Ref. 2).
 SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 90.7%; Score 78; DB 1; Length 536;
 Best Local Similarity 85.7%; Pred. No. 0.00095;
 Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
 DB 287 TWHYYLNGRVATK 300

RESULT 4
 HPSE_MOUSE STANDARD; PRT; 535 AA.
 ID HPSE_MOUSE
 AC Q6YGL; O8K3K3;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 11-OCT-2005, sequence version 2.
 DT 07-MAR-2006, entry version 13.
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
 GN Name:Hpse; Synonyms:Hpa;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RC STRAIN=SUL/J; TISSUE=Spleen;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
 RA Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
 RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND
 RP SUBUNITS.
 RC STRAIN=FVB; TISSUE=Embryo;
 RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
 RA Mao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
 RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
 RT "Cloning, expression, and purification of mouse heparanase.";
 RL Protein Expr. Purif. 26:425-431(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
 RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,
 RA Li J.-P.;
 RT "Processing of macromolecular heparin by heparanase.";
 RL J. Biol. Chem. 278:35152-35158(2003).
 RN [4]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].

RC STRAIN=C57BL/6J, and MOD; TISSUE=Thymus;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Ainslie V., Allen J.E.,
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 RA Di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustigich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
 RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugitara K., Sultana R., Takenaka Y., Taki K.,
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Tesdale R.D., Liu E.T., Brusic V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kawashina T., Kojima M., Kondo S., Konno H., Nakano K., Nimomiya N.,
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Wataniki A., Okamura-Ono Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome.";
 RL Science 309:1559-1563(2005).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Contributes to metastasis and angiogenesis
 CC (By similarity).
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and
 CC magnesium (By similarity). Inhibited by laminarin sulfate and, to
 CC a lower extent, by heparin and sulfamin.
 CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 5;
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products.
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium
 CC (By similarity).
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
 CC -----
 CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms

CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR ENBL; AF359507; AAQ15188.1; -; mRNA.
 DR ENBL; AY077467; AAL76083.1; -; mRNA.
 DR ENBL; AY151051; AAN41636.1; -; mRNA.
 DR ENBL; AK040471; BAC30600.1; -; mRNA.
 DR ENBL; AK154628; BAE32725.1; -; mRNA.
 DR Ensembl; ENSMUSG00000035273; Mus musculus.
 DR MGI; MGI:1343124; Hpse.
 DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 KW Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
 KW Magnesium; Membrane; Signal. By similarity.
 FT SIGNAL 1 27
 FT CHAIN 28 101
 FT PROPEP 102 149
 FT CHAIN 150 535
 FT REGION 150 154
 FT REGION 262 272
 FT ACT_SITE 217 217
 FT ACT_SITE 335 335
 FT CARBOHYD 154 154
 FT CARBOHYD 192 192
 FT CARBOHYD 209 209
 FT CARBOHYD 230 230
 FT CARBOHYD 451 451
 FT CONFLICT 206 206
 FT CONFLICT 212 212
 FT CONFLICT 230 232
 FT CONFLICT 335 335
 FT CONFLICT 342 342
 FT CONFLICT 455 455
 FT CONFLICT 531 531
 FT CONFLICT 535 535
 FT SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;
 Query Match 89.5%; Score 77; DB 1; Length 535;
 Best Local Similarity 85.7%; Pred. No. 0.0014;
 Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 TWHYYLNGRTATR 14
 Db 286 TWHYYLNGRTATK 299
 RESULT 5
 Q333X6 SPAJD
 ID Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
 AC Q333X6;
 DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
 DT 06-DEC-2005, sequence version 1.
 DT 07-FEB-2006, entry version 3.
 DE Heparanase.
 GN Name=hpa;
 OS Spalax judaei (Blind subterranean mole rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Spalacidae; Spalacinae; Spalax.
 OX NCBI_TaxID=134510;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 RL cloning and identification of a novel splice variant";
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviivi A.;

```
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338
|||||
|

RESULT 6
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]_TaxID=164324;
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FD19DCBBD99DE CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338
|||||
|

RESULT 7
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1]_TaxID=191382;
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
```

```
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48BEFEC7D0BCB34 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338
|||||
|

RESULT 8
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1]_TaxID=164323;
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338
|||||
|

RESULT 9
HPSE_BOVIN
```

```

ID HPSE_BOVIN STANDARD; PRT; 545 AA.
AC Q9MYO;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2006, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (BC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
GN Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;
RZ Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -1- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -1- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
DR EMBL; AF281160; AAF87301.2; -; mRNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW SIGNAL.
FT SIGNAL 1 37 By similarity.
FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
FT FT /FTID=PRO_0000042256.
FT PROPEP 112 159 Linker peptide.
FT FT /FTID=PRO_0000042257.
FT CHAIN 160 545 Heparanase 50 kDa subunit (By
FT similarity).
FT FT /FTID=PRO_0000042258.
FT REGION 160 164 Heparin/HS-binding (Potential).
FT REGION 272 282 Heparin/HS-binding (Potential).
FT ACT_SITE 227 227 Proton donor (Potential).
FT ACT_SITE 345 345 Nucleophile (Potential).
FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;
Query Match 86.0%; Score 74; DB 1; Length 545;
Best Local Similarity 78.6%; Pred. NO. 0.0042;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TWHYYLNGRTATR 14
Db 296 TWHYYVNGRIATK 309
RESULT 10
O4TGC8_TETNG PRELIMINARY; PRT; 255 AA.
ID O4TGC8_TETNG
AC Q4TGC8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00001168001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=1549614; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Aubourg S., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaut S., Bafnau L., Fisher S., Luthalla G., Dessat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
DR EMBL; CAAB01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1 255
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;
Query Match 76.7%; Score 66; DB 2; Length 255;
Best Local Similarity 76.9%; Pred. NO. 0.036;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TWHYYLNGRTAT 13
Db 130 TWHYYLDGREAS 142
RESULT 11
O4SYF6_TETNG PRELIMINARY; PRT; 533 AA.
ID O4SYF6_TETNG

```

AC	Q4SYF6;
AD	19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT	19-JUL-2005, sequence version 1.
DT	07-FEB-2006, entry version 4.
DE	Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE	(Fragment).
GN	ORFNames=GSTENG00010356001;
OS	Tetraodon nigroviridis (Green puffer);
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC	Tetraodontidae; Tetraodontidae; Tetraodon.
OX	NCBI_TaxID=99883;
OX	[1]
RP	NUCLEOTIDE SEQUENCE.
RP	PubMed=15496914; DOI=10.1038/nature03025;
RX	Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA	Maucell E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA	Nicaud S., Jaffe D., Fisher S., Lutfailla G., Dossat C., Segurens B.,
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA	Anthouard V., Jubin C., Castelli V., Katkina M., Vacherie B.,
RA	Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA	Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,
RA	Farra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA	Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA	Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT	the early vertebrate proto-karyotype.";
RL	Nature 431:946-957(2004).
RL	[2]
RP	NUCLEOTIDE SEQUENCE.
RG	Genoscope; Whitehead Institute Centre for Genome Research;
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
-1	CAUTION: The sequence shown here is derived from an
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
CC	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
CC	EMBL; CAAE01012073; CAF94326.1; -; Genomic_DNA.
DR	NON TER 1
FT	NON TER 533 533
FT	SEQUENCE 533 AA; 50100 MW; 9B00A7C8780100FF CRC64;
SQ	
Query Match	76.7%; Score 66; DB 2; Length 533;
Best Local Similarity	76.9%; Pred. No. 0.076;
Matches 10;	Conservative 2; Mismatches 1; Indels 0; Gaps
Qy	1 TWHYYLNGRTAT 13
	: :
Db	250 TWHYYLDGREA 262
RESULT 12	
ID	QBT108_BOMMO PRELIMINARY; PRT; 515 AA.
AC	QBT108;
DT	01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT	01-JUN-2002, sequence version 1.
DT	07-FEB-2006, entry version 10.
DE	Heparanase-like protein.
DE	Name=Bmhepa;
OS	Bombyx mori (Silk moth).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC	Bombycidae; Bombyx.
OX	NCBI_TaxID=7091;
OX	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=p50; TISSUE=Posterior silkgland;

ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAP1780, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG0003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Athouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,
RA Ceraud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Kallis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; CAEE01007180; CAP89852.1; -; Genomic_DNA.
FT NON TER 597 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
Query Match 60.5%; Score 52; DB 2; Length 597;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TWHHYLNGR 10
Db 334 TWQHYWDGR 343
|||:|:|:|
|||:|:|:|
RESULT 15
OS9462 PYRHO PRELIMINARY; PRT; 1179 AA.
AC OS9462;
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.
DT 01-AUG-1998, sequence version 1.
DT 07-FEB-2006, entry version 30.
DE 1179aa long hypothetical chromosome assembly protein.
GN OrderedLocustNames=PH1798;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194; DOI=10.1093/dnares/5.2.55;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

Yamamoto S., Sekine M., Baba S., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Mauchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; BA000001; BAA30917.1; -; Genomic_DNA.
DR PIR; F71190; F71190.
DR HSPF; Q9X0R4; 1E69.
DR SMR; OS9462; 2-167, 1012-1170.
DR BioCyc; PHOR53953:PH1798-MONOMER; -.
DR GO; GO:0005694; C:chromosome; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0005525; F:GTP binding; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0051277; P:chromosome organization and biogenesis (sen. .; IEA.
DR InterPro; IPR003439; ABC transp_like.
DR InterPro; IPR005289; GTP_bd.
DR InterPro; IPR010935; SMC hinge.
DR InterPro; IPR003395; SMC N.
DR InterPro; IPR011891; SMC prok A.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF06470; SMC hinge; 1.
DR Pfam; PF02463; SMC_N; 2.
DR ProDom; PD000006; ABC transporter; 1.
DR TIGRFAMs; TIGR00650; MG442; 1.
DR TIGRFAMs; TIGR02169; SMC prok A; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1179 AA; 135657 MW; 24A61BESF3864493 CRC64;
Query Match 59.3%; Score 51; DB 2; Length 1179;
Best Local Similarity 81.8%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 4 HYVINGRATR 14
Db 113 HYVINGRATR 123
|||:|:|:|
|||:|:|:|
RESULT 16
Q4WRV8 ASPFU PRELIMINARY; PRT; 124 AA.
AC Q4WRV8;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-MAR-2006, entry version 5.
DE Hypothetical protein.
GN ORFNames=Afuig14930;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Af293 / CBS 101355 / FGSC A1100;
RX PubMed=16372009; DOI=10.1038/nature04332;
RA Nierman W.C., Pain A., Anderson M.J., Wortman J.R., Kim H.S.,
RA Arroyo J., Berrian M., Abe K., Archer D.B., Bermejo C., Bennett J.W.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N.D., Feldblyum T.V., Fischer R.,
RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.J.,
RA Haas H., Harris D.E., Horiuchi H., Huang J., Humphray S., Jimenez J.,
RA Keller N., Kfour H., Kitamoto K., Kobayashi T., Konzack S.,
RA Kulkarni R., Kumagai T., Laifon A., Latge J.-P., Li W., Lord A.,
RA Lu C., Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M.,

RA Monod M., Mouyna I., Mulligan S., Murphy L.D., O'Neil S., Paulsen I.,
RA Penalba M.A., Perteira M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitz E., Rawlins N., Rajandream M.A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J.R., Yu J.-H., Fraser C.M., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B.G., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT *Aspergillus fumigatus*.";
RL Nature 438:1151-1156(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAHF01000004; EAL90824.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 124 AA; 14143 MW; DF32481969BDB831 CRC64;

Query Match 58.1%; Score 50; DB 2; Length 124;
Best Local Similarity 57.1%; Pred. No. 6.1;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 TWHVYLNQRTATR 14
|||||
DB 64 TWHGYLGGQVRER 77

RESULT 17
ID Q2KXXB_BORAV PRELIMINARY; PRT; 52 AA.
AC Q2KXXB;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Hypothetical protein.
GN ORFNames=BAV2400;
OS Bordetella avium 197N.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=360910;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=197N;
RA Sebaihia M.;
RT "The genome sequence of the poultry pathogen *Bordetella avium*, and
RT genomic comparisons with related species infecting mammals.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AM167904; CAJ50010.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 52 AA; 6063 MW; 2C7181A0BCAACAA3 CRC64;

Query Match 57.0%; Score 49; DB 2; Length 52;
Best Local Similarity 53.8%; Pred. No. 3.6;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHHYVYLNQRTATR 14
|||||
DB 13 WHHWLNLTRVNTX 25

RESULT 18
ID Q7P1C9_CHRVO PRELIMINARY; PRT; 138 AA.
AC Q7P1C9;

DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN OrderedLocNames=CV0284; ORFNames=CV_0284;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=536;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA Alves-Gomes J.A., Andrade E.M., Aratipe J., de Azaújo M.F.F.,
RA Astolfi-Pilho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
RA Bordignon J., Brígido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Chueire L.M.O.,
RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.P.,
RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,
RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunarez H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of *Chromobacterium violaceum* reveals
RT remarkable and exploitable bacterial adaptability.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AE016825; AA057963.1; -; Genomic DNA.
DR Biocyc; CV10243365:CV0284-MONOMER; -;
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 138 AA; 15009 MW; BEC7720F3E6500CF CRC64;

Query Match 54.7%; Score 47; DB 2; Length 138;
Best Local Similarity 70.0%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WHHYVYLNQRT 11
|||||
DB 45 WHGYVYHGRT 54

RESULT 19
ID Q5C4P3_SCHJA PRELIMINARY; PRT; 215 AA.
AC Q5C4P3;
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 2.
DT 07-FEB-2006, entry version 6.
DE SJCHGC07775 protein (Fragment).
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatoidea; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP NUCLEOTIDE SEQUENCE.

RA Liu F., Lu J., Hu W., Wang S.-Y., Cui S.-J., Chi M., Yan Q.,
 RA Wang X.-R., Song H.-D., Xu X.-N., Wang J.-J., Zhang X.-L., Wang Z.-Q.,
 RA Xue C.-L., Brindley P.J., McManus D.P., Yang P.-Y., Feng Z., Chen Z.,
 RA Han Z.-G.;
 RT "New Perspectives on Host-parasite Interplay by Comparative
 RT Transcriptional and Proteomic Analyses of the Human Blood Fluke,
 RT *Schistosoma japonicum*,"
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL: AY809492; AAX25381.2; -: mRNA.
 FT NON_TER 1 215
 FT NON_TER 215 215
 SQ SEQUENCE 215 AA; 25120 MW; D02C3167C2B80C3 CRC64;
 Query Match 54.7%; Score 47; DB 2; Length 215;
 Best Local Similarity 63.6%; Pred. No. 32;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 WHYYLNGRTA 12
 DB || |||: ||:
 85 WHPYLSNRTS 95
 RESULT 20
 Q7VJ45 HELHP PRELIMINARY; PRT; 382 AA.
 AC Q7VJ45;
 DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.
 DT 01-OCT-2003, sequence version 1.
 DT 07-FEB-2006, entry version 17.
 DE Citrate synthase PrpC (EC 2.3.3.1).
 GN Names:prpC; OrderedLocusNames:HH0398; ORFNames:HH_0398;
 OS Helicobacter hepaticus.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=32025;
 [1]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=ATCC 51449 / 381;
 RX MEDLINE=22709201; PubMed=12810954; DOI=10.1073/pnas.1332093100;
 RA Suerbaum S., Josenhans C., Stenzenbach T., Drescher B., Brandt P.,
 RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,
 RA Holland R., Klein K., Koening J., Macko L., Mendz G.L., Nyakatura G.,
 RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
 RT "The complete genome sequence of the carcinogenic bacterium
 RT *Helicobacter hepaticus*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL: AE017125; AAP76995.1; -: Genomic_DNA.
 DR HSSP; O34002; IA59.
 DR BioCy; HHPF235279:HH0398-MONOMER; -.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0008415; F:acyltransferase activity; IEA.
 DR GO; GO:0004108; F:citrate (S)-synthase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006092; P:main pathways of carbohydrate metabolism; IEA.
 DR InterPro; IPR011278; Cit_synth_II.
 DR InterPro; IPR002020; Citrate synth.
 DR PANTHER; PTHR11739; Citrate synth; 1.
 DR Pfam; PF00285; Citrate synth; 1.
 DR PRINTS; PRO0143; CITRISYNTHASE.
 DR TIGRFAMs; TIGR01800; cit_synth_II; 1.
 DR Acyltransferase; Complete proteome; Transferase.
 SQ SEQUENCE 382 AA; 42813 MW; C51D9727E13656D3 CRC64;
 Query Match 54.7%; Score 47; DB 2; Length 382;
 Best Local Similarity 66.7%; Pred. No. 57;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 WHYYLNGR 10
 DB ||||: ||:
 143 WHYYHHNGK 151
 RESULT 21
 Q41029 GIBZE PRELIMINARY; PRT; 383 AA.
 ID Q41029 GIBZE
 AC Q41029;
 DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
 DT 16-AUG-2005, sequence version 1.
 DT 07-FEB-2006, entry version 4.
 DE Hypothetical protein.
 DE Hypothetical protein.
 OS ORFNames=FG09109.1;
 GN Gibberella zeae (Fusarium graminearum).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
 OX NCBI_TaxID=5518;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=PH-1 / NRRL 31084;
 RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
 RA Boukhgelter B., Butler J., Calvo S.E., Camarata J., Chang J.,
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearlano K.,
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
 RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,
 RA Galagan J.E., Gardyna S., Gherre S., Graham L., Grand-Pierre N.,
 RA Hafez N., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,
 RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,
 RA Karatas A., Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G.,
 RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,
 RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,
 RA Meneus L., Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C.,
 RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,
 RA O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,
 RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,
 RA Rogov P., Roman J., Schauer S., Schupback R., Seaman S., Severy P.,
 RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,
 RA Stubbs M., Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
 RA Lander E.S.;
 RL "Fusarium graminearum genome sequence,"
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL: AACM01000370; EAA78159.1; -: Genomic_DNA.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 383 AA; 42750 MW; 993AC5392560831A CRC64;
 Query Match 54.7%; Score 47; DB 2; Length 383;
 Best Local Similarity 72.7%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 HHYYLNGRTAT 13
 DB ||||: |||:
 272 HHYYFAGRPAT 282
 RESULT 22
 KCMA1 CANFA STANDARD; PRT; 1159 AA.
 ID KCMA1 CANFA
 AC Q28265;

RESULT 23
 Q8BSE9_MOUSE PRELIMINARY; PRT; 64 AA.
 AC Q8BSE9;
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 18.
 DE 12 days embryo embryonic body between diaphragm region and neck cDNA,
 DE product: siatyltransferase 5, full insert sequence. (Fragment).
 GN Name: Stgal2; Synonyms: Siat5;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and
 RC neck;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Methods Enzymol. 303:19-44 (1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and
 RC neck;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama R., Ravasi F., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 RA Ambesi-Impombato A., Apeeler R., Aturaliya R.N., Bailey T.L.,
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.B., Dalla E., Dall'Amico B.P., de Bono B., Della Gatta G.,
 RA di Bernardo C.F., Down T., Engstrom P., Fagiolini M., Faulkner G.,
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Hummelbeck L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
 RA Petrovsky N., Piazza S., Reed J.C., Ring B.Z., Ringwald M.,
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
 RA Tammojä K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome.";
 RL Science 309:1559-1563 (2005).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and

RC neck;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG RIKEN Genome Exploration Research Group, and Genome Science Group
 RT (Genome Network Core Team) and the FANTOM Consortium;
 RL "Antisense Transcription in the Mammalian Transcriptome.";
 RL Science 309:1564-1566 (2005).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and
 RC neck;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.A.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi L., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Saito R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Sultana R., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuen Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and
 RC neck;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690 (2001).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and
 RC neck;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to


```

Matches      8;  Conservative      1;  Mismatches      3;  Indels      0;  Gaps      0;

Qy      3 HHYLYNGRTATR 14
|      ||||| :||
Db      154 HRYLYNGNSAYR 165

RESULT 26
QZULB2 ASPOR
ID      QZULB2 ASPOR      PRELIMINARY;      PRT;      208 AA.
AC      QZULB2;
DT      24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT      24-JAN-2006, sequence version 1.
DT      07-MAR-2006, entry version 3.
DE      N-acetyltransferase.
GN      ORFNames=AO0900030000513;
OS      Aspergillus oryzae.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX      NCBI_TaxID=5062;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=RIB 40;
RX      PubMed=16372010; DOI=10.1038/nature04300;
RA      Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA      Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA      Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA      Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
RA      Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA      Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA      Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA      Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA      Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA      Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA      Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA      Kuhara S., Ogasawara N., Kikuchi H.;
RT      "Genome sequencing and analysis of Aspergillus oryzae.";
RL      Nature 438:1157-1161(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      -----
CC      EMBL; AP007155; BAB57683.1; -; Genomic DNA.
DR      GO; GO:0016740; F:transferase activity; IEA.
KW      Transferase.
SQ      SEQUENCE      208 AA;      23330 MW;      EEB1E4BFA8F469F6 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 208;
Best Local Similarity      66.7%;      Pred. No. 45;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

Qy      3 HHYLYNGRTATR 14
|      ||||| :||
Db      155 HRYLYNGNSAYR 166

RESULT 27
QZS4E2 NEUCR
ID      QZS4E2 NEUCR      PRELIMINARY;      PRT;      213 AA.
AC      QZS4E2;
DT      15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT      15-DEC-2003, sequence version 1.
DT      07-FEB-2006, entry version 10.
DE      Hypothetical protein.
GN      ORFNames=NCU02417.1;
OS      Neurospora crassa.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX      NCBI_TaxID=5141;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=74-OR23-1A / FGSC 987;
RX      MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;

Query Match      53.5%;      Score 46;      DB 2;      Length 208;
Best Local Similarity      66.7%;      Pred. No. 45;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

Qy      3 HHYLYNGRTATR 14
|      ||||| :||
Db      155 HRYLYNGNSAYR 166

RESULT 27
QZS4E2 NEUCR
ID      QZS4E2 NEUCR      PRELIMINARY;      PRT;      213 AA.
AC      QZS4E2;
DT      15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT      15-DEC-2003, sequence version 1.
DT      07-FEB-2006, entry version 10.
DE      Hypothetical protein.
GN      ORFNames=NCU02417.1;
OS      Neurospora crassa.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX      NCBI_TaxID=5141;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=74-OR23-1A / FGSC 987;
RX      MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;

```

```

RA      Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA      Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA      Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA      Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,
RA      Werner-Washburne M., Belltremlnikoff C.P., Kinsey J.A., Braun E.L.,
RA      Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
RA      Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA      Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M.,
RA      Mauceli E., Bielek C., Rudd S., Frishman D., Krystofova S.,
RA      Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA      Macino G., Catchside D.E.A., Li W., Pratt R.J., Osmani S.A.,
RA      DeSouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA      Yarden O., Plamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA      Natvig D.O., Alex L.A., Mannheim G., Ebbole D.J., Freitag M.,
RA      Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.W.;
RL      "The genome sequence of the filamentous fungus Neurospora crassa.";
RL      Nature 422:859-868(2003).
CC      -!- CAUTION: The sequence shown here is derived from an
CC      EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC      preliminary data.
CC      -----
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      -----
CC      EMBL; AABX01000368; EAA30363.1; -; Genomic DNA.
DR      GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR      InterPro; IPR000182; GCN5acetyl_trans.
KW      Pfam; PF00583; Acetyltransf_1; 1.
KW      Hypothetical protein.
SQ      SEQUENCE      213 AA;      24470 MW;      75FF1BF4E5AA4B39 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 213;
Best Local Similarity      66.7%;      Pred. No. 46;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

Qy      3 HHYLYNGRTATR 14
|      ||||| :||
Db      176 HRYLYNGNSAYR 187

RESULT 28
Q5B387 EMENI
ID      Q5B387 EMENI      PRELIMINARY;      PRT;      213 AA.
AC      Q5B387;
DT      26-APR-2005, integrated into UniProtKB/TrEMBL.
DT      26-APR-2005, sequence version 1.
DT      07-MAR-2006, entry version 6.
DE      Hypothetical protein.
GN      ORFNames=AN4993.2;
OS      Aspergillus nidulans FGSC A4.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiales; Trichocomaceae; Emericella.
OX      NCBI_TaxID=227321;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=FGSC 4;
RX      PubMed=16372000; DOI=10.1038/nature04341;
RA      Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R.,
RA      Batzoglou S., Lee S.-I., Bastuerkmen M., Spevak C.C., Clutterbuck J.,
RA      Kapitonov V., Jurka J., Sczocchio C., Farman M., Butler J.,
RA      Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C.,
RA      Bouchier C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S.,
RA      Doonan J.H., Yu J., Vienken K., Pain A., Freitag M., Selker E.U.,
RA      Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T.,
RA      Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W.,
RA      Caddick M., Hynes M., Paolletti M., Fischer R., Miller B.L., Dyer P.S.,
RA      Sachs M.S., Osmani S.A., Birren B.W.;
RT      "Sequencing of Aspergillus nidulans and comparative analysis with A.
RT      fumigatus and A. oryzae.";
RL      Nature 438:1105-1115(2005).
CC      -!- CAUTION: The sequence shown here is derived from an
CC      EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
CC      preliminary data.

```

```
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AACD0100084; EAA61071.1; -: Genomic DNA.
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR InterPro; IPR000182; GCN5acetyl_trans.
DR Pfam; PF00593; Acetyltransf_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 213 AA; 24228 MW; 8BDB1FBADD83BCBD CRC64;

Query Match 53.5%; Score 46; DB 2; Length 213;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHVYLNGRTATR 14
| | | | | : | |
Db 156 HRYLNGNSAYR 167

RESULT 29
Q4WXL6 ASPFU PRELIMINARY; PRT; 279 AA.
AC Q4WXL6;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Acetyltransferase, GNAT family, putative.
DE ORFNames=Atu3g09940;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=AF293 / CBS 101355 / FGSC A1100;
RX PubMed=16372009; DOI=10.1038/nature04332;
RA Nierman W.C., Pain A., Anderson M.J., Wortman J.R., Kim H.S.,
RA Arroyo J., Berriman M., Abe K., Archer D.B., Bernejo C., Bennett J.W.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman N., Fedorova N., Fedorova N.D., Feldblyum T.V., Fischer R.,
RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.J.,
RA Haas H., Harris D.E., Horiuchi H., Huang J., Humphray S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Konzack S.,
RA Kulkarni R., Kumagai T., Lafont A., Latge J.-P., Li W., Lord A.,
RA Lu C., Majeros W.H., May G.S., Miller B.L., Mohammed Y., Molina M.,
RA Monod M., Mouyna I., Mulligan S., Murphy L.D., O'Neil S., Paulsen I.,
RA Penalva M.A., Perlea M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitsch E., Rawlins N., Rajandream M.A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J.R., Yu J.-H., Fraser C.M., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B.G., Denning D.W.;
RA "Genomic sequence of the pathogenic and allergenic filamentous fungus
RA Aspergillus fumigatus.";
RT Nature 438:1151-1156(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; RAHF0100002; EAL92587.1; -: Genomic DNA.
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
KW Complete proteome; Transferase.
SQ SEQUENCE 279 AA; 31318 MW; 193FB4FD8EA2407C CRC64;

Query Match 53.5%; Score 46; DB 2; Length 279;
```

```
Best Local Similarity 66.7%; Pred. No. 60;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHVYLNGRTATR 14
| | | | | : | |
Db 231 HRYLNGNSAYR 242

RESULT 30
Q5DZ88 VIBF1 PRELIMINARY; PRT; 301 AA.
AC Q5DZ88;
DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.
DT 15-MAR-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
DE OrderedLocusNames=VFA0838;
OS Vibrio fischeri (Strain ATCC 700601 / ES114).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=312309;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP Ruby E.G., Urbanowski M., Campbell J., Dunn A., Faini M., Gunsalus R.,
RX PubMed=15703294; DOI=10.1073/pnas.0409900102;
RA Lostroh P., Lupp C., McCann J., Millikan D., Schaefer A., Stabb E.,
RA Stevens A., Visick K., Whistler C., Greenberg E.P.;
RT "Complete genome sequence of Vibrio fischeri: a symbiotic bacterium
RT with pathogenic congeners.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3004-3009(2005).
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000021; AAW87908.1; -: Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 301 AA; 34109 MW; D85E45C8469A9693 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 301;
Best Local Similarity 61.5%; Pred. No. 65;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTAT 13
| | | | | : | |
Db 55 TPXHYTFNGKAT 67

RESULT 31
Q8BSA0 MOUSE PRELIMINARY; PRT; 349 AA.
AC Q8BSA0;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE 12 days embryo embryonic body between diaphragm region and neck cDNA,
DE RIKEN full-length enriched library, clone:9430050G04
DE product:sialyltransferase 5, full insert sequence.
DE Name=St3gal2; Synonyms=Stat5;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

NUCLEOTIDE SEQUENCE.
RP TISSUE=Embryonic body between diaphragm region and neck;
RC STRAIN=C57BL/6J;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6875(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]

NUCLEOTIDE SEQUENCE.
```


RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crowe M.B., Dalla E., Dalrymple B.P., De Bono B., Della Gatta G., Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruen Y., Salzberg S.L., Sandelin A., Schneider C., Schonbach C., Sekiguchi K., Semple C.A., Seno S., Seosa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K., Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yeghi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Inamori K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;
 RL "The transcriptional landscape of the mammalian genome."; Science 309:1559-1563(2005).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;
 RG "Antisense Transcription in the Mammalian Transcriptome."; Science 309:1564-1566(2005).
 RL [4]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru N., Hasegawa Y., Nogami H., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L., Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.D., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,

RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirokawa T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).
 RL [5]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., De Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto Y., Storch K.-F., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S., Hayashizaki Y.;
 RL "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Akizawa J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).
 RL [8]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;
 RX PubMed=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y


```

RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Tova T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK034863; BAC28859.1; -; mRNA.
DR Ensembl; ENSMUSG00000031749; Mus musculus.
DR MGI; MGI:99427; St3gal2.
DR GO; GO:0030173; C:integral to Golgi membrane; IEA.
DR GO; GO:0008373; F:sialyltransferase activity; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro; IPR001675; Glyco_trans_29.
DR InterPro; IPR012163; Sialyl_trans.
DR Pfam; PF00777; Glyco_transf_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 349 AA; 40033 MW; 372926AE6379186E CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

RESULT 32
Q5ZJJO.CHICK PRELIMINARY; PRT; 349 AA.
AC Q5ZJJO;
DT 23-NOV-2004, integrated into UniProtKB/TREMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN ORFNames=RCJMB04.17119;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Biagodataski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ720444; CAG32103.1; -; mRNA.
DR GO; GO:0030173; C:integral to Golgi membrane; IEA.
DR GO; GO:0008373; F:sialyltransferase activity; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro; IPR001675; Glyco_trans_29.
DR InterPro; IPR012163; Sialyl_trans.
DR Pfam; PF00777; Glyco_transf_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Hypothetical protein.
SQ SEQUENCE 349 AA; 39804 MW; 2F80876451449E32 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

```

```

RESULT 33
Q70D58.CHICK PRELIMINARY; PRT; 349 AA.
AC Q70D58;
DT 05-JUL-2004, integrated into UniProtKB/TREMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Alpha2,3-sialyltransferase.
GN Name=ST3GAL-II;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC PubMed=15843597; DOI=10.1093/glycob/cwi063;
RA Harduin-Leipers A., Mollicone R., Delannoy P., Oriol R.;
RT "The animal sialyltransferases and sialyltransferase-related genes: a
RT phylogenetic approach.";
RL Glycobiology 15:805-817(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ585761; CAE51385.2; -; mRNA.
DR GO; GO:0030173; C:integral to Golgi membrane; IEA.
DR GO; GO:0008373; F:sialyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro; IPR001675; Glyco_trans_29.
DR InterPro; IPR012163; Sialyl_trans.
DR Pfam; PF00777; Glyco_transf_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Transferase.
SQ SEQUENCE 349 AA; 39683 MW; B14F4CEB464F0456 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

```

```

RESULT 34
SIA4B.HUMAN STANDARD; PRT; 350 AA.
AC Q16842; O00654;
DT 11-JUL-2002, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 49.
DE CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase
DE (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)
DE (ST3GAL4.2) (SIAT4-B) (ST3Gal II).
GN Name=ST3GAL2; Synonyms=SIAT4B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Liver;
RX MEDLINE=97079181; PubMed=8920913; DOI=10.1006/bbrc.1996.1660;

```

Kim Y.-J., Kim K.-S., Kim S.-H., Kim C.-H., Ko J.H., Choe I.-S., Tsuji S., Lee Y.-C.;
RT "Molecular cloning and expression of human Gal beta 1.3GalNAc alpha
RT 2,3-sialyltransferase [hST3Gal II]."
RL Biochem. Biophys. Res. Commun. 228:324-327(1996).
RN [2]
RN
RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RP MEDLINE=97409982; PubMed=9266697;
RX Giordanengo V., Lefebvre J.C.;
RT "Cloning and expression of cDNA for a human Gal beta1-3GalNAc
RT alpha2,3-sialyltransferase from the CEM T cell line.";
RL Eur. J. Biochem. 247:558-566(1997).
RN [3]
RN
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RP TTSUSE=Lung;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Sraueberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
RA Altshul S.F., Jordan H., Buotow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting M., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.C., Krzywicki M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC
CC -1- FUNCTION: It may be responsible for the synthesis of the sequence
CC NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal
CC carbohydrate groups of certain glycoproteins, oligosaccharides and
CC glycolipids, STAA4 and SIAT4B sialylate the same acceptor
CC substrates but exhibit different Km values.
CC -1- CATALYTIC ACTIVITY: CMP-N-acetylneuraminate + beta-D-galactosyl-
CC 1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-
CC acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC galactosaminyl-R.
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC form in trans cisternae of Golgi. Soluble form in body fluids (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and heart
CC and to a much lesser extent in brain, placenta, liver and
CC pancreas. Scarcely detectable in lung and kidney.
CC -1- PTM: The soluble form derives from the membrane form by
CC proteolytic processing [By similarity].
CC -1- SIMILARITY: Belongs to the glycosyltransferase 29 family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
CC ENBL; U63090; AAB40389.1; -; mRNA.
CC ENBL; X96667; CAA65447.1; -; mRNA.
CC ENBL; BC036777; AAH36777.1; -; mRNA.
CC PIR; JC5251; JC5251.
CC Ensembl; ENSG00000157350; Homo sapiens.
CC HGNC; HGNC:10863; ST3GAL2.
CC MIM; 607188; gene.
CC GO; GO:00003836; F:beta-galactoside alpha-2,3-sialyltransferas. .; TAS.
CC GO; GO:0006040; P:amino sugar metabolism; TAS.
CC GO; GO:0006464; P:protein modification; TAS.
CC InterPro; IPR012163; Sialyl trans.
CC InterPro; IPR012163; Sialyl trans.
CC Pfam; PF00777; Glyco trans 29; 1.
CC PIRSF; PIRSF005557; Sialyl trans; 1.
CC

KW	Glycoprotein; Glycosyltransferase; Golgi stack; Membrane;
KW	Signal-anchor; Transferase; Transmembrane.
FT	CHAIN 1 350
FT	CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase.
FT	/FTid=PRO 0000149258.
FT	Cytoplasmic (Potential).
FT	Signal-anchor for type II membrane protein (Potential).
FT	Luminal (Potential).
FT	TOPO DOM 28 350
FT	CARBOHYD 92 92
FT	CARBOHYD 211 211
FT	N-linked (GlcNAc . .) (Potential).
FT	DISULFID 152 291
FT	By similarity.
SQ	SEQUENCE 350 AA; 40173 MW; EYE40CF26D9CB725 CRC64;
Query Match	
Best Local Similarity 53.5%; Score 46; DB 1; Length 350;	
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps	
QY	2 WHHYLNGRTA 12
DB	307 WHHYWENRYA 317
RESULT 35	
SLA4B PANTR	STANDARD; PRT; 350 AA.
AC	Q6XB58;
DT	15-FEB-2005, integrated into UniProtKB/Swiss-Prot.
DT	05-JUL-2004, sequence version 1.
DT	07-FEB-2006, entry version 15.
DE	CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase) (ST3GalA-2) (SIAT4-B) (ST3Gal II).
GN	Name=ST3GAL2; Synonyms=SIAT4B;
OS	Pan troglodytes (Chimpanzee).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.
OC	NCBI_TaxID=9598;
RN	[1]
RP	NUCLEOTIDE SEQUENCE [MRNA].
RA	Harduin-Lepeers A., Martinez-Duncker I., Mollicone R., Delannoy P., Oriol R.;
RT	"Phylogeny of sialyltransferases.";
RL	Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC	FUNCTION: It may be responsible for the synthesis of the sequence NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal carbohydrate groups of certain glycoproteins, oligosaccharides and glycolipids. SIAT4A and SIAT4B sialylate the same acceptor substrates but exhibit different Km values (By similarity).
CC	CATALYTIC ACTIVITY: CMP-N-acetylneuraminase + beta-D-galactosyl-1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-acetylneuraminyl-1,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-galactosaminyl-R.
CC	PATHWAY: Glycosylation.
CC	SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound form in trans cisternae of Golgi. Soluble form in body fluids (By similarity).
CC	PTM: The soluble form derives from the membrane form by proteolytic processing (By similarity).
CC	SIMILARITY: Belongs to the glycosyltransferase 29 family.
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
DR	EMBL; AJ744804; CAG32840.1; -; mRNA.
DR	InterPro; IPR001675; Glyco trans 29.
DR	InterPro; IPR012163; Sialyl trans.
DR	Pfam; PF00777; Glyco transf 29.1.
DR	PIRSF; PIRSF005557; Sialyl trans; 1.
KW	Glycoprotein; Glycosyltransferase; Golgi stack; Membrane

KW Signal-anchor; Transferrase; Transmembrane.
 FT CHAIN 1 350 CMP-N-acetylneuraminase-beta-
 FT galactosamide-alpha-2,3-
 FT sialyltransferase.
 FT /FTId=PRO.0000149260.
 FT TOPO_DOM 1 6 Cytoplasmic (Potential).
 FT TRANSMEM 7 27 Signal-anchor for type II membrane
 FT protein (Potential).
 FT TOPO_DOM 28 350 Luminal (Potential).
 FT CARBOHYD 211 211 N-linked (GlcNAc...) (Potential).
 FT DISULFID 152 291 By similarity.
 SQ SEQUENCE 350 AA; 40074 MW; FCE9932A2D9CB73A CRC64;
 Query Match 53.5%; Score 46; DB 1; Length 350;
 Best Local Similarity 63.6%; Pred. No. 76;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WHYYLNGRTA 12
 Db 307 WHYYENNRYA 317
 RESULT 36
 STIA4B RAT STANDARD; PRT; 350 AA.
 AC Q11205;
 DT 01-OCT-1996, integrated into UniProtKB/Swiss-Prot.
 DT 01-OCT-1996, sequence version 1.
 DT 07-FEB-2006, entry version 40.
 DE CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase
 DE (EC 2.4.99.-) (Beta-galactosidase alpha-2,3-sialyltransferase) (Alpha
 DE 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)
 DE (ST3GALA.2) (SIAT4-B) (ST3Gal II).
 GN Name=St3gal2; Synonyms=SIAT4B, Siat5;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RC TISSUE=Brain;
 RX MEDLINE=94193584; PubMed=8144500;
 RA Lee Y.-C., Kojima N., Wada E., Kurosawa N., Nakaoka T., Hamamoto T.,
 RA Tsuji S.;
 RT "Cloning and expression of cDNA for a new type of Gal beta 1,3GalNAc
 RT alpha 2,3-sialyltransferase.";
 RL J. Biol. Chem. 269:10028-10033(1994).
 CC -!- FUNCTION: It may be responsible for the synthesis of the sequence
 CC NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal
 CC carbohydrate groups of certain glycoproteins, oligosaccharides and
 CC glycolipids. SIAT4A and SIAT4B sialylate the same acceptor
 CC substrates but exhibit different Km values.
 CC -!- CATALYTIC ACTIVITY: CMP-N-acetylneuraminyl- + beta-D-galactosyl-
 CC 1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-acetyl-
 CC acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
 CC galactosaminyl-R.
 CC -!- PATHWAY: Glycosylation.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
 CC form in trans cisternae of Golgi. Soluble form in body fluids.
 CC -!- PTM: The soluble form derives from the membrane form by
 CC proteolytic processing.
 CC -!- SIMILARITY: Belongs to the glycosyltransferase 29 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL; X76988; CAA54293.1; -; mRNA.
 CC PIR; B54420; B54420.
 CC Ensembl; ENSRNOG00000017932; Rattus norvegicus.
 CC RGD; 68413; Siat4b.
 CC InterPro; IPR001675; Glyco trans_29.
 CC Intron; IPR012163; Sialyl_trans.

DR Pfam; PF00777; Glyco transf_29; 1.
 DR PIRSF; PIRSF005557; Sialyl_trans; 1.
 KW Glycoprotein; Glycosyltransferase; Golgi stack; Membrane;
 KW Signal-anchor; Transferrase; Transmembrane.
 FT CHAIN 1 350 CMP-N-acetylneuraminase-beta-
 FT galactosamide-alpha-2,3-
 FT sialyltransferase.
 FT /FTId=PRO.0000149261.
 FT TOPO_DOM 1 6 Cytoplasmic (Potential).
 FT TRANSMEM 7 27 Signal-anchor for type II membrane
 FT protein (Potential).
 FT TOPO_DOM 28 350 Luminal (Potential).
 FT CARBOHYD 211 211 N-linked (GlcNAc...) (Potential).
 FT DISULFID 152 291 By similarity.
 SQ SEQUENCE 350 AA; 40166 MW; 87E6494F02D0BE1 CRC64;
 Query Match 53.5%; Score 46; DB 1; Length 350;
 Best Local Similarity 63.6%; Pred. No. 76;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WHYYLNGRTA 12
 Db 307 WHYYENNRYA 317
 RESULT 37
 Q6H8M9 BOVIN PRELIMINARY; PRT; 350 AA.
 ID Q6H8M9 BOVIN PRELIMINARY; PRT; 350 AA.
 AC Q6H8M9;
 DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
 DT 19-JUL-2004, sequence version 1.
 DT 07-MAR-2006, entry version 14.
 DE Sialyltransferase ST3Gal-II.
 GN Name=SIAT4B; Synonyms=SIAT4B;
 GN Bos taurus (Bovine).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 OC Pecora; Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX PubMed=15843597; DOI=10.1093/glycob/cwi063;
 RA Harduin-Lepera A., Mollicone R., Delannoy P., Oriol R.;
 RT "The animal sialyltransferases and sialyltransferase-related genes: a
 RT phylogenetic approach.";
 RL Glycobiology 15:805-817(2005).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ll Hereford; TISSUE=Ascending colon;
 RA Moore S., Alexander L., Brownstein M., Guan L., Lobo S., Meng Y.,
 RA Taniguchi M., Wang Z., Yu J., Prange C., Schreiber K., Shenmen C.,
 RA Wagner L., Bala M., Barbazuk S., Barber S., Babakaiff R., Beland J.,
 RA Chun E., Del Rio L., Gibson S., Hanson R., Kirkpatrick R., Liu J.,
 RA Matsuo C., Mayo M., Santos R.R., Stott J., Tsai M., Wong D.,
 RA Siddiqui A., Holt R., Jones S.J., Marra M.A.;
 RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL; AJ748841; CAG44450.1; -; mRNA.
 CC EMBL; BC105252; AAI05253.1; -; mRNA.
 CC Ensembl; ENSBTAG0000002006; Bos taurus.
 CC GO; GO:0030173; C:integral to Golgi membrane; IEA.
 CC GO; GO:0008373; F:sialyltransferase activity; IEA.
 CC GO; GO:0006486; P:protein amino acid glycosylation; IEA.
 CC InterPro; IPR01675; Glyco trans_29.
 CC Intron; IPR012163; Sialyl_trans.
 CC Pfam; PF00777; Glyco transf_29; 1.
 CC PIRSF; PIRSF005557; Sialyl_trans; 1.
 KW Glycosyltransferase; Transferrase.
 SQ SEQUENCE 350 AA; 40009 MW; 9B6853BB8A761063 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 350;
 Best Local Similarity 63.6%; Pred No. 76;
 Matches 7; Conservativity 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
 |||||
 DB 307 WHYWNRYA 317

RESULT 38
 Q8BPLO MOUSE
 ID Q8BPLO MOUSE PRELIMINARY; PRT; 350 AA.
 AC Q8BPLO;
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
 DT 01-MAR-2003, sequence version 1.
 DT 07-FEB-2006, entry version 22.
 DE O clone neonate eyeball cDNA, RIKEN full-length enriched library.
 DE O clone neonate eyeball cDNA, RIKEN full-length enriched library.
 DE O clone neonate eyeball cDNA, RIKEN full-length enriched library.
 DE (S73 beta-galactoside alpha-2,3-sialyltransferase 2) (Adult male
 DE cerebellum cDNA, RIKEN full-length enriched library, clone:1520403117
 DE product:sialyltransferase 5, full insert sequence).
 GN Name-St3gal2; Synonyms-Siat5;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RT "High-efficiency full-length cDNA cloning.";
 RL Methods Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RC Carninci P., Katayama T., Katayama S., Gough J., Frith M.C., Maeda N.,
 Ouyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
 Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
 Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
 Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni F., Pesole G.,
 Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
 Rost B., Ruan Y., Salzberg S.L., Sanderlin A., Schneider C.,
 Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 Shibata Y., Shimada K., Shimada K., Silva D., Sinclair B.,
 Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
 Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 Grimmond S.M., Teasdale R.D., Liu E.T., Brusica V., Quackenbush J.,
 Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,

RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
 Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome.";
 RL Science 309:1559-1563(2005).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG. RIKEN Genome Exploration Research Group, and Genome Science Group
 (Genome Network Core Team) and the FANTOM Consortium;
 RT "Antisense Transcription in the Mammalian Transcriptome.";
 RL Science 309:1564-1566(2005).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuta M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 Blake J.A., Bradt D., Brusica V., Chothia C., Corbani L.E., Cousins S.,
 Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 Shiraki T., Waki K., Waki J., Aizawa K., Arakawa T., Fukuda S.,
 Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberti P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
 Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;


```
RESULT 40
Q6NU41 XENLA
ID Q6NU41 XENLA PRELIMINARY; PRT; 351 AA.
AC Q6NU41;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 13.
DE MGC81265 protein.
GN Name=MGC81265;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC068760; AAH68760.1; -; mRNA.
DR GO; GO:0030173; C:integral to Golgi membrane; IEA.
DR GO; GO:0008373; F:sialyltransferase activity; IEA.
DR GO; GO:0006486; P:protein amino acid glycosylation; IEA.
DR InterPro; IPR01675; Glyco trans 29.
DR InterPro; IPR012163; Sialyl trans.
DR Pfam; PF00777; Glyco trans_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
SQ SEQUENCE 351 AA; 40217 MW; 4A1AF6DD38400DD36 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 351;
Best Local Similarity 63.6%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHHYLNGRTA 12
| | | | |
DB 308 WHHYWENRYA 318
```

```
RESULT 41
Q4AXU3 9BURK
ID Q4AXU3 9BURK PRELIMINARY; PRT; 892 AA.
AC Q4AXU3;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Transposase tm3.
GN ORFNames=BprODRAFT_4286;
OS Polaromonas sp. JS666.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Polaromonas.
OX NCBI_TaxID=236591;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Polaromonas sp.
RT JS666."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft assembly of Polaromonas sp. JS666."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAF02000007; EAM38976.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004803; F:transposase activity; IEA.
DR GO; GO:0006313; P:DNA transposition; IEA.
SQ SEQUENCE 892 AA; 101919 MW; 14E1EFDBA791B848 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 892;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHHYLNGRTATR 14
| | | | |
DB 863 WRHVHLNGRYAFR 875

RESULT 42
O80438 ARATH
ID O80438 ARATH PRELIMINARY; PRT; 190 AA.
AC O80438;
DT 01-NOV-1998, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1998, sequence version 1.
DT 07-FEB-2006, entry version 23.
DB Putative acetyltransferase (At2g38130).
GN OrderedLocusNames=At2g38130;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
```

```
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Roundley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN NUCLEOTIDE SEQUENCE.
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RN NUCLEOTIDE SEQUENCE.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RN NUCLEOTIDE SEQUENCE.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AC003028; AAC27162.1; -; Genomic_DNA.
DR EMBL; BT010841; AAR24208.1; -; mRNA.
DR EMBL; BT012615; AAT06434.1; -; mRNA.
DR PIR; T01245; T01245.
DR TAIR; At2g38130; -.
DR GO; GO:0008080; F.N-acetyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000182; GCN5acetyl trans.
DR Pfam; PF00593; Acetyltransf_1; 1.
KW Transferase.
SQ SEQUENCE 190 AA; 22123 MW; 124BD9D94D0A491 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 190;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHVYLNRTATR 14
Db 147 YHYLNGMDAFR 158

RESULT 43
Q2UR20 ASPOR
ID Q2UR20 ASPOR PRELIMINARY; PRT; 448 AA.
AC Q2UR20;
DT 24-JAN-2006, integrated into UniprotKB/TREMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-MAR-2006, entry version 3.
DE Predicted protein.
GN ORFNames=AO090005000633;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RC PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagaoka H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA Kuhara S., Ogasawara N., Kikuchi H.,
```

```
RT "Genome sequencing and analysis of Aspergillus oryzae.";
RL Nature 438:1157-1161(2005).
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By
CC similarity).
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007151; BAB55675.1; -; Genomic_DNA.
DR Membrane; Transmembrane; Transport.
SQ SEQUENCE 448 AA; 49144 MW; 740F5AB85B3C9790 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 WHHYL 7
Db 184 WHHYL 189

RESULT 44
Q33X5 SPAJD
ID Q33X5 SPAJD PRELIMINARY; PRT; 558 AA.
AC Q33X5;
DT 06-DEC-2005, integrated into UniprotKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085494; CAJ30021.1; -; mRNA.
SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F5849EE7 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 558;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 YLNGRTATR 14
Db 313 YLNGRIATK 322

RESULT 45
Q4RHG9 TETNG
ID Q4RHG9 TETNG PRELIMINARY; PRT; 612 AA.
AC Q4RHG9;
DT 19-JUL-2005, integrated into UniprotKB/TREMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Chromosome 3 SCAF15050, whole genome shotgun sequence. (Fragment).
```

GN ORFNames=GSTENG0034356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Blemont C., Skalli Z., Bottolier L., Poullain J., De Berardinis V.,
RA Craumat C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC
CC EMBL; CAE01015050; CAG12163.1; -; Genomic DNA.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0008152; P: metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR PRINTS; PF00501; AMP-binding; 1.
DR PROSITE; PS00154; AMPBINDING.
FT NON_TER 1 1
FT NON_TER 612 612
SQ SEQUENCE 612 AA; 67282 MW; BC8AGE6727D0E297 CRC64;
Query Match 52.3%; Score 45; DB 2; Length 612;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 TWHHYLYNGRTATR 14
Db 9 TWKYFYIAARTAKR 22
RESULT 46
Q6PY34_9POTY PRELIMINARY; PRT; 3050 AA.
AC Q6PY34;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Polypeptide.
OS Hordeum mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Rymovirus.
OX NCBI_TaxID=41764;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC PV81;

RX PubMed=15193921; DOI=10.1016/j.virol.2004.03.014;
RA Stenger D.C., French R.;
RT "Functional replacement of Wheat streak mosaic virus HC-Pro with the
RT corresponding citron from a diverse array of viruses in the family
RT Potyviridae";
RL Virology 323:257-267(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC PV81;
RX PubMed=1549142; DOI=10.1007/s00705-004-0396-6;
RA French R., Stenger D.C.;
RT "Genome sequences of Agropyron mosaic virus and Hordeum mosaic virus
RT support reciprocal monophyly of the genera Potyvirus and Rymovirus in
RT the family Potyviridae";
RL Arch. Virol. 150:299-312(2005).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC
CC EMBL; AY623627; AAS65455.2; -; Genomic RNA.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0005524; F: ATP binding; IEA.
DR GO; GO:0008026; F: ATP-dependent helicase activity; IEA.
DR GO; GO:0004137; F: cysteine-type endopeptidase activity; IEA.
DR GO; GO:0003723; F: RNA binding; IEA.
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006508; P: proteolysis; IEA.
DR GO; GO:0019079; P: transcription; IEA.
DR InterPro; IPR001410; DEAD/DEAH_N.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR002540; Pept_S30_PotY_P1.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001456; Peptidase_C6.
DR InterPro; IPR001592; Poty_coat.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007095; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_P3D.
DR Pfam; PF00270; DEAD_1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00851; Peptidase_C6; 1.
DR Pfam; PF01577; Peptidase_S30; 1.
DR Pfam; PF00767; Poty_coat; 1.
DR Pfam; PF00680; RDRP_1; 1.
DR PRINTS; PR00966; NIAPOTYPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Polyprotein.
SQ SEQUENCE 3050 AA; 344716 MW; 455C2C163256F888 CRC64;
Query Match 51.7%; Score 44.5; DB 2; Length 3050;
Best Local Similarity 69.2%; Pred. No. 1.2e+03;
Matches 9; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 2 WHHYLYN-GRAT 13
Db 41 WHAYLYNAGRELT 53
RESULT 47
Q7V6T4_PROMM
ID Q7V6T4_PROMM PRELIMINARY; PRT; 161 AA.
AC Q7V6T4;
DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Putative ribosomal-protein-alanine acetyltransferase (EC 2.3.1.128).
GN OrderedLocusNames=PMT1060;
OS Prochlorococcus marinus (strain MIT 9313).

OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=74547;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
RA Rocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chain P.,
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation."
RL Nature 424:1042-1047(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BX572098; CAE21235.1; -; Genomic_DNA.
DR BioCyc; PMAR74547; PMT1060-MONOMER; -;
DR GO; GO:0008415; F:acetyltransferase activity; IEA.
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR GO; GO:0008999; F:ribosomal-protein-alanine N-acetyltransferase; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000182; GCN5acetyl trans.
DR Pfam; PF00583; Acetyltransf 1; 1.
KW Acyltransferase; Complete proteome; Transferase.
SQ SEQUENCE 161 AA; 17972 MW; D18ED45E9E0C4493 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 161;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHYINGRTA 12
||| |||
Db 135 HHYINGQDA 144

RESULT 48
QASB93_TETNG
ID QASB93_TETNG PRELIMINARY; PRT; 234 AA.
AC QASB93;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Chromosome undetermined SCAF14676, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0021075001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lufalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionet C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype."
RL Nature 431:946-957(2004).
RN [2]

RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -! SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAE01014676; CAG02089.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_lambrepresr.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON TER 1
SQ SEQUENCE 234 AA; 26319 MW; E0B2315DC2EC02BB CRC64;

Query Match 51.2%; Score 44; DB 2; Length 234;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNG 9
|:|:| |
Db 50 TYHHYNLNG 58

RESULT 49
Q404U7_9RHOB
ID Q404U7_9RHOB PRELIMINARY; PRT; 236 AA.
AC Q404U7;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Peptidase M50.
GN ORFNames=JannDRAFT_4278;
OS Jannaschia sp. CCS1.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Jannaschia.
OX NCBI_TaxID=290400;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCS1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Jannaschia sp. CCS1."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CCS1;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Jannaschia sp. CCS1."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AAIG01000018; EAM65398.1; -; Genomic_DNA.

DR GO: GO:0004222; P:metalloendopeptidase activity; IEA.
DR GO: GO:0006508; P:proteolysis; IEA.
DR InterPro: IPR008915; Peptidase_M50.
DR Pfam: PF02163; Peptidase_M50; I.
SQ SEQUENCE 236 AA; 25641 MW; 348E8A2AD1B64F72 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 236;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 TWHHYLNQRTAT 14
:|:|:|:|:|
Db 177 SWLHRFVNGATNR 190

RESULT 50
Q5W6T7 ORYSA
ID Q5W6T7 ORYSA PRELIMINARY; PRT; 260 AA.
AC Q5W6T7
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein P0018A03.12.
GN Name=P0018A03.12;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
RA Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,
RA Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L.,
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
RA Wu H.-P., Shaw J.-F., Yu Y., Rambo T., Currie J., Collura K.,
RA Soderlund C., Wing R.;
RT "Oryza sativa PAC P0018A03 genomic sequence."
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AC134346; AAV44183.1; -; Genomic DNA.
DR Gramene; Q5W6T7; -;
DR InterPro: IPR006458; DUF623_pln.
DR Pfam: PF04844; DUF623; 1.
DR TIGRFAMs: TIGR01568; A_thal_3678; 1.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 28936 MW; A23DB19E7D40C568 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 260;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHVYLNQRTAT 13
|||||:|
Db 56 HHVYLNQRTAT 66

Search completed: June 5, 2006, 12:53:58
Job time : 159.74 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 104.384 Seconds

(without alignments)
65.702 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPGKKVVLGETSSAY 15

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2599679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2599679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : A_Geneseq_8.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

9: geneseqp2005s.*

10: geneseqp2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	ADL16423	Adl16423 Human hep
2	82	100.0	15	ADR88215	Adr88215 Human hep
3	82	100.0	15	ADT78182	Adt78182 Functiona
4	82	100.0	15	ADU71218	Adu71218 Human hep
5	82	100.0	15	AEA42431	Aea42431 Human hep
6	82	100.0	386	ADR88207	Adr88207 Human mat
7	82	100.0	386	ADT78174	Adt78174 45kDa sub
8	82	100.0	386	ADY27057	Ady27057 Heparanas
9	82	100.0	386	ADZ18995	Adz18995 Human hep
10	82	100.0	386	AEA42423	Aea42423 Human mat
11	82	100.0	460	ADY27061	Ady27061 Heparanas
12	82	100.0	486	AE887589	Aeb87589 Human hep
13	82	100.0	492	ADZ18996	Adz18996 Hep106 co
14	82	100.0	493	AE887562	Aeb87562 Human hep
15	82	100.0	495	ADZ18999	Adz18999 Hep109 co
16	82	100.0	497	AE887587	Aeb87587 Human hep
17	82	100.0	501	ADZ19000	Adz19000 HepG3 co
18	82	100.0	507	ADZ19005	Adz19005 HepG6 co
19	82	100.0	508	ADY27058	Ady27058 Human ina
20	82	100.0	526	ADZ19006	Adz19006 Hepyalur
21	82	100.0	527	ABW07815	Abw07815 Chicken s
22	82	100.0	527	ABW02018	Abw02018 Chimeric
23	82	100.0	527	ADO63825	Ado63825 Chimeric

24	82	100.0	527	9	ADZ19004	Adz19004 HepG4 co
25	82	100.0	530	2	AAV34173	Aav34173 Human pre
26	82	100.0	532	2	AAV17083	Aav17083 Seq ID No
27	82	100.0	543	2	AAV02345	Aav02345 A human h
28	82	100.0	543	2	AAV17082	Aav17082 Human hep
29	82	100.0	543	3	AAV57590	Aav57590 Human hep
30	82	100.0	543	3	AAV08849	Aav08849 Amino aci
31	82	100.0	543	3	AAV52990	Aav52990 Human hep
32	82	100.0	543	4	AAV97635	Aav97635 Human hep
33	82	100.0	543	4	AAV86206	Aav86206 Human hep
34	82	100.0	543	5	ABB07813	Abb07813 Human hep
35	82	100.0	543	7	ADD18950	Add18950 Human dis
36	82	100.0	543	7	ADG88800	Adg88800 Human hpa
37	82	100.0	543	8	ADL16379	Adl16379 Human hep
38	82	100.0	543	8	ADK52086	Adk52086 Human ato
39	82	100.0	543	8	ADM48716	Adm48716 Human hpa
40	82	100.0	543	8	ADM48759	Adm48759 Human hpa
41	82	100.0	543	8	ADN05074	Adn05074 Antipsori
42	82	100.0	543	8	ADN04902	Adn04902 Antipsori
43	82	100.0	543	8	ADO63831	Ado63831 Human hep
44	82	100.0	543	8	ADO63832	Ado63832 Human hep
45	82	100.0	543	8	ADO63822	Ado63822 Human hep
46	82	100.0	543	8	ADR80372	Adr80372 Heparanas
47	82	100.0	543	8	ADR88210	Adr88210 Human pre
48	82	100.0	543	8	ADP25079	Adp25079 PRO poly
49	82	100.0	543	8	ADT78177	Adt78177 Human hep
50	82	100.0	543	9	ADY27036	Ady27036 Human hep
51	82	100.0	543	9	AEA42466	Aea42466 Human hep
52	82	100.0	543	9	AEA42426	Aea42426 Human hep
53	82	100.0	543	10	AE896848	Aee896848 Human hep
54	82	100.0	545	6	ABP56822	Abp56822 Human hep
55	82	100.0	545	7	ADR16012	Adr16012 G-coupled
56	82	100.0	545	8	ADL93951	Adl93951 Human G-c
57	82	100.0	556	9	ADZ19010	Adz19010 Heparanas
58	82	100.0	570	9	ADZ19008	Adz19008 Heparanas
59	82	100.0	588	2	AAV30124	Aav30124 A human p
60	82	100.0	592	2	AAV02346	Aav02346 A human h
61	82	100.0	592	3	AA808850	Aab88850 Amino aci
62	82	100.0	592	7	ADG88804	Adg88804 Human SK-
63	82	100.0	592	8	ADL16383	Adl16383 Human hep
64	82	100.0	592	8	ADM48720	Adm48720 Human SK-
65	82	100.0	592	9	AEA42461	Aea42461 Human hep
66	77	93.9	15	9	ADU71069	Adu71069 Human hep
67	77	93.9	19	9	ADO63828	Ado63828 Human hep
68	77	93.9	380	2	AAV17085	Aav17085 Rat hepar
69	77	93.9	380	2	AAV17084	Aav17084 Mouse hep
70	77	93.9	535	3	AB808851	Aab08851 A murine
71	77	93.9	535	5	AB807811	Abb07811 Mouse hep
72	77	93.9	535	7	ADG88834	Adg88834 Mouse hpa
73	77	93.9	535	8	ADL16413	Adl16413 Mouse hep
74	77	93.9	535	8	ADM48750	Adm48750 Mouse hpa
75	77	93.9	535	8	ADR88208	Adr88208 Mouse hep
76	77	93.9	535	8	ADT78175	Adt78175 Mouse hep
77	77	93.9	535	9	ADY27033	Ady27033 Murine he
78	77	93.9	535	9	AEA42424	Aea42424 Mouse hep
79	77	93.9	536	5	ABB07812	Abb07812 Rat hepar
80	77	93.9	536	8	ADR88209	Adr88209 Rat hepar
81	77	93.9	536	8	ADT78176	Adt78176 Rat hepar
82	77	93.9	536	9	AEA42425	Aea42425 Rat hepar
83	77	93.9	536	9	AEA42425	Aea42425 Rat hepar
84	77	93.9	543	4	AA888361	Aab88361 Human mem
85	77	93.9	543	9	ADY63087	Ady63087 Human clo
86	76	92.7	527	8	ADO63827	Ado63827 Chimeric
87	76	92.7	527	8	ADO63826	Ado63826 Chimeric
88	76	92.7	543	8	ADO63824	Ado63824 Human hep
89	76	92.7	543	8	ADO63823	Ado63823 Human hep
90	75	91.5	15	9	ADU71068	Adu71068 Human h
91	73	89.0	523	5	ABB07814	Abb07814 Chicken h
92	73	89.0	523	7	ABW02017	Abw02017 Chicken h
93	73	89.0	523	8	ADR88211	Adr88211 Chicken h
94	73	89.0	523	8	ADT78178	Adt78178 Chicken h
95	73	89.0	523	9	ADY27037	Ady27037 Chicken h
96	73	89.0	523	9	AEA42427	Aea42427 Chicken h

97	70	85.4	15	9	ADU70873	Adu70873 Human hep	170	42	51.2	833	8	ADM72219	Human TAS
98	68	82.9	545	9	ADY27034	Ady27034 Bovine hep	171	42	51.2	833	8	ADP24693	PRO polyp
99	64	78.0	15	9	ADU70843	Adu70843 Human hep	172	42	51.2	836	8	Adx68344	Plant ful
100	63	76.0	15	9	ADU71217	Adu71217 Human hep	173	42	51.2	847	8	Adr39751	Human kin
101	59	72.0	15	9	ADU70974	Adu70974 Human hep	174	42	51.2	1027	9	Ady70305	Human bet
102	54	65.9	13	2	AAV34185	Aav34185 Human pre	175	42	51.2	1366	2	Aar71702	Collagen
103	53	64.6	9	9	ADU70701	Adu70701 Human hep	176	42	51.2	1366	3	Aay56801	Human pre
104	53	64.6	15	9	ADU71291	Adu71291 Human hep	177	42	51.2	1366	3	Aay96123	Collagen
105	51	62.2	9	9	ADU70549	Adu70549 Human hep	178	42	51.2	1366	4	Abb50293	Collagen
106	51	62.2	488	4	AAAB31471	Aab31471 Amino aci	179	42	51.2	1366	5	AAE16476	Human col
107	51	62.2	488	4	AAAB31469	Aab31469 Amino aci	180	42	51.2	1366	5	ABB90751	Human Tum
108	50	61.0	9	9	ADU70474	Adu70474 Human hep	181	42	51.2	1366	5	ABB90766	Human Tum
109	49	59.8	9	9	ADU70739	Adu70739 Human hep	182	42	51.2	1366	5	ABB90741	Human Tum
110	48	58.3	9	9	ADU70550	Adu70550 Human hep	183	42	51.2	1366	5	ABB80734	Protein s
111	47	57.5	9	9	ADU70473	Adu70473 Human hep	184	42	51.2	1366	5	ABB09626	Amino aci
112	47	57.3	9	9	ADU70418	Adu70418 Human hep	185	42	51.2	1366	6	ABU54448	Human tum
113	47	57.3	15	9	ADU70973	Adu70973 Human hep	186	42	51.2	1366	6	ABU54473	Human tum
114	47	57.3	488	4	AAAB31470	Aab31470 Amino aci	187	42	51.2	1366	6	ABU54458	Human tum
115	46	56.1	561	6	ABU27696	Abu27696 Protein e	188	42	51.2	1366	6	ABR92065	Human cer
116	46	56.1	639	4	ABG21490	Abg21490 Novel hum	189	42	51.2	1366	7	ADF13076	Human pro
117	46	56.1	1180	6	ABU28742	Abu28742 Protein e	190	42	51.2	1366	8	ADK70437	Respirato
118	45	54.9	134	2	AAAR72393	Aar72393 Agmenellu	191	42	51.2	1366	8	ADQ29669	Human col
119	45	54.9	177	8	ADS73853	Ads73853 A. thalia	192	42	51.2	1366	8	ADR16801	Human col
120	45	54.9	523	6	ABU45111	Abu45111 Protein e	193	42	51.2	1366	8	ADR16426	Human col
121	44	53.7	87	4	ABG29684	Abg29684 Novel hum	194	42	51.2	1366	8	ADR99147	Collagen,
122	44	53.7	93	5	ADK36889	Adk36889 Novel hum	195	42	51.2	1366	9	ADV87001	Collagen
123	44	53.7	104	6	AAU48166	Aau48166 Propionib	196	42	51.2	1366	9	ADV70234	Tumor-ass
124	44	53.7	104	6	ABMA4685	Abm44685 Propionib	197	42	51.2	1366	9	ADZ70510	Human pro
125	44	53.7	177	4	ABG24421	Abg24421 Novel hum	198	42	51.2	1366	9	ADZ70624	Human pro
126	44	53.7	336	6	ADB23092	Adb23092 Environme	199	42	51.2	1366	9	AEA04491	Human pro
127	44	53.7	341	8	ADQ65272	Adq65272 Novel hum	200	42	51.2	1366	9	AED74603	Human pla
128	44	53.7	354	4	ABG10413	Abg10413 Novel hum	201	42	51.2	1366	10	AEF69981	Colorecta
129	44	53.7	488	4	AAAB31472	Aab31472 Amino aci	202	41	50.0	68	2	AAV31469	N. tabacu
130	44	53.7	563	6	AAU34625	Aau34625 B. coli c	203	41	50.0	68	3	AAAB25828	AP2 domai
131	44	53.7	563	6	ABU28684	Abu28684 Protein e	204	41	50.0	91	5	AAM50384	Rat hepar
132	44	53.7	1167	6	ABU27927	Abu27927 Protein e	205	41	50.0	168	4	AAM58491	Propionib
133	43	52.4	9	9	ADU70581	Adu70581 Human hep	206	41	50.0	168	6	ABM55010	Propionib
134	43	52.4	15	9	ADU70872	Adu70872 Human hep	207	41	50.0	172	4	AAU65751	Propionib
135	43	52.4	349	4	AAU50487	Aau50487 Propionib	208	41	50.0	172	6	ABM62270	Propionib
136	43	52.4	349	6	ABMA47006	Abm47006 Propionib	209	41	50.0	192	8	ADX91766	Plant ful
137	42	51.2	58	4	AAAO07208	Aao07208 Human pol	210	41	50.0	222	5	AMX87478	Plant ful
138	42	51.2	65	4	AAU32257	Aau32257 Novel hum	211	41	50.0	225	5	AMX47870	Tobacco p
139	42	51.2	83	4	AAO09648	Aao09648 Human pol	212	41	50.0	225	7	ADD15139	Tobacco e
140	42	51.2	114	4	AAO13397	Aao13397 Human pol	213	41	50.0	266	10	AEF11596	Soybean m
141	42	51.2	265	7	AAE38636	Aae38636 Human col	214	41	50.0	273	8	ADX65843	Plant ful
142	42	51.2	283	3	AAAG44159	Aag44159 Arabidops	215	41	50.0	281	8	ADT58907	Plant pol
143	42	51.2	283	3	AAAG43083	Aag43083 Arabidops	216	41	50.0	281	10	AEF29235	Lead_Cere
144	42	51.2	305	4	AAAG73640	Aag73640 Human col	217	41	50.0	282	8	ADX70963	Plant ful
145	42	51.2	340	9	ADW17545	Adw17545 Pinus rad	218	41	50.0	304	8	ADX68795	Plant ful
146	42	51.2	341	3	AAAG44158	Aag44158 Arabidops	219	41	50.0	312	10	AEF11656	Rice meth
147	42	51.2	341	3	AAAG43082	Aag43082 Arabidops	220	41	50.0	326	10	AEF11652	Rice meth
148	42	51.2	341	8	ADS73843	Ads73843 A. thalia	221	41	50.0	331	5	AAM50383	Human hep
149	42	51.2	341	9	AED60915	Aed60915 Thale cre	222	41	50.0	332	10	AEF11638	Wheat met
150	42	51.2	341	10	AEF26570	Aef26570 A. thalia	223	41	50.0	349	10	AEF11642	Wheat met
151	42	51.2	354	9	ABU17546	Abu17546 Pinus rad	224	41	50.0	355	8	ADX80019	Plant ful
152	42	51.2	355	6	ABU70801	Abu70801 Human adi	225	41	50.0	355	8	ADX80019	Plant ful
153	42	51.2	366	6	ABU15516	Abu15516 Protein e	226	41	50.0	365	5	ABG60934	Novel flo
154	42	51.2	385	5	ADK35029	Adk35029 Novel hum	227	41	50.0	365	5	ABG60934	Rice meth
155	42	51.2	385	7	ABO78366	AbO78366 Pseudomon	228	41	50.0	379	6	ABR40913	Deducd a
156	42	51.2	386	6	ABR41621	AbR41621 Human DIT	229	41	50.0	382	10	AEF11562	Soybean m
157	42	51.2	397	7	ADJ68198	Adj68198 Human hea	230	41	50.0	384	10	AEF11630	Soybean v
158	42	51.2	464	9	ADW17551	Adw17551 Pinus rad	231	41	50.0	384	10	AEF11630	Soybean m
159	42	51.2	487	7	ABO62870	AbO62870 Klebsiell	232	41	50.0	392	10	AEF11560	Soybean m
160	42	51.2	495	7	ADB37566	AdB37566 Neural th	233	41	50.0	439	4	AUO7423	Human hep
161	42	51.2	495	7	ADJ69822	AdJ69822 Human hea	234	41	50.0	445	7	ADF59424	Human hep
162	42	51.2	534	2	AAW12844	Aaw12844 Pro-alpha	235	41	50.0	470	5	AAE18328	Human hep
163	42	51.2	535	2	AAW12841	Aaw12841 Truncated	236	41	50.0	475	8	ADS23751	Bacterial
164	42	51.2	623	2	AAW12843	Aaw12843 Pro-alpha	237	41	50.0	480	4	AAV97634	Human hep
165	42	51.2	674	3	AAAB53439	Aab53439 Human col	238	41	50.0	480	4	AAU07418	Novel hum
166	42	51.2	740	8	ADR87614	Adr87614 Human Typ	239	41	50.0	480	4	AAAB85217	Heparanas
167	42	51.2	772	4	ABG01438	Abg01438 Novel hum	240	41	50.0	492	4	AAAB84664	Amino aci
168	42	51.2	809	2	AAV29672	Aav29672 Human cer	241	41	50.0	528	5	AAE18327	Human hep
169	42	51.2	833	6	AAO30330	Aao30330 Human MAP	242	41	50.0	534	4	AAAB85216	Heparanas

243	41	50.0	534	5	ABP69310	Abp69310 Human pol	316	40	48.8	4928	2	AAV39300	Aay39300 Spnd a po
244	41	50.0	534	5	AAW50337	Aam50337 Human pre	317	40	48.8	4933	6	ABP57681	Abp57681 Saccharop
245	41	50.0	538	4	AAV97633	Aay97633 Human hep	318	40	48.8	5588	2	AAV39301	Aay39301 Spne a po
246	41	50.0	582	5	AAE18326	Aae18326 Human hep	319	40	48.8	5588	4	AAB70969	Aab70969 S. spinos
247	41	50.0	592	4	AAV97632	Aay97632 Human hep	320	39.5	48.2	171	9	AED71520	Aed71520 Corynebact
248	41	50.0	592	4	AAAB1062	Aab81062 Human hep	321	39.5	48.2	383	4	AAG93067	Aag93067 C glutami
249	41	50.0	592	4	AAU07424	Aau07424 Human hep	322	39.5	48.2	385	7	ABO65564	Abg65564 Klebsiell
250	41	50.0	592	4	AAAB85215	Aab85215 Heparanas	323	39	47.6	9	ADU70769	Adu70769 Human hep	
251	41	50.0	1130	8	ADT89111	Adt89111 Soybean M	324	39	47.6	15	ADU71070	Adu71070 Human hep	
252	40	48.8	63	4	AAW23316	Aam23316 Human EST	325	39	47.6	30	ADX83892	Adx83892 Bacillus	
253	40	48.8	63	4	AAW23351	Aam23351 Human EST	326	39	47.6	62	ADU88825	Adx88825 Ovule dev	
254	40	48.8	68	2	AAV31473	Aay31473 A. thalia	327	39	47.6	84	AAO09797	Aao09797 Human pol	
255	40	48.8	68	3	ABW25832	Abw25832 AP2 domai	328	39	47.6	91	1	AAP61040	Aap61040 N-termina
256	40	48.8	69	2	AAV31470	Aay31470 N. tabacu	329	39	47.6	100	3	ABAB33003	Abab33003 Pinus rad
257	40	48.8	69	3	ABW25829	Abw25829 AP2 domai	330	39	47.6	103	4	AAO12779	Aao12779 Human pol
258	40	48.8	103	7	ABO63688	Aboc63688 Klebsiell	331	39	47.6	107	5	ABP11307	Abp11307 Human ORF
259	40	48.8	109	3	AAV19715	Aag19715 Arabidops	332	39	47.6	125	4	AAO12263	Aao12263 Human pol
260	40	48.8	112	3	AAV19714	Aag19714 Arabidops	333	39	47.6	127	9	AEA26323	Aea26323 Stress to
261	40	48.8	119	4	AAU02524	Aau02524 Anti-adip	334	39	47.6	128	9	AEA27246	Aea27246 Stress to
262	40	48.8	151	9	ADW18623	Adw18623 Eucalyptu	335	39	47.6	128	10	AEF28690	Aef28690 Lead Cere
263	40	48.8	167	3	ABP32789	Aab32789 Eucalyptu	336	39	47.6	130	4	AAW17187	Aam17187 Peptide #
264	40	48.8	193	10	Aef11612	Aef11612 Soybean m	337	39	47.6	130	4	ABB36190	Abb36190 Peptide #
265	40	48.8	193	10	Aef11602	Aef11602 Soybean m	338	39	47.6	130	4	AAW29682	Aam29682 Peptide #
266	40	48.8	198	8	ADW70784	Adx70784 Plant ful	339	39	47.6	130	4	ABW30994	Abw30994 Peptide #
267	40	48.8	200	7	ADD30468	Adi30468 Plant yie	340	39	47.6	130	4	ABB21565	Abb21565 Protein #
268	40	48.8	200	8	ADI43757	Adi43757 Plant tra	341	39	47.6	130	4	AAW69355	Aam69355 Human bon
269	40	48.8	204	8	ADW70856	Adx70856 Plant ful	342	39	47.6	130	4	AAW56968	Aam56968 Human bra
270	40	48.8	210	4	ABG13427	Abg13427 Novel hum	343	39	47.6	130	4	ABW51031	Abw51031 Human liv
271	40	48.8	210	8	ADW70847	Adx70847 Plant ful	344	39	47.6	130	4	AAW04884	Aam04884 Peptide #
272	40	48.8	213	4	AAU54490	Aau54490 Propionib	345	39	47.6	130	5	ABG38974	Abg38974 Human pep
273	40	48.8	213	6	ABW51009	Abw51009 Propionib	346	39	47.6	130	8	ADW63744	Adw63744 Transcript
274	40	48.8	220	8	ADW71560	Adx71560 Plant ful	347	39	47.6	130	9	AEA26237	Aea26237 Stress to
275	40	48.8	222	5	AAU93108	Aau93108 Arabidops	348	39	47.6	130	9	AEA27248	Aea27248 Stress to
276	40	48.8	222	7	ADD55844	Adi55844 Thalecres	349	39	47.6	156	2	AAW10439	Aaw10439 Prepro su
277	40	48.8	222	8	ADW1595	Adw1595 Thalecres	350	39	47.6	158	8	ADW66520	Adw66520 Plant ful
278	40	48.8	222	8	ADW73293	Adw73293 Thale cre	351	39	47.6	166	7	ABW86570	Abw86570 Rice abio
279	40	48.8	222	9	ADW17092	Adw17092 Eucalyptu	352	39	47.6	166	10	Aef28702	Aef28702 Lead Cere
280	40	48.8	229	7	ADC84613	Adc84613 Chandipur	353	39	47.6	171	4	AAW78376	Aam78376 Human pro
281	40	48.8	232	4	AAE08544	Aae08544 Phage M10	354	39	47.6	172	4	AAW00037	Aam00037 Beta-1, 3
282	40	48.8	232	8	ADW65815	Adw65815 Plant ful	355	39	47.6	173	4	ABW04008	Abw04008 Novel hum
283	40	48.8	261	7	ABO75261	Abw75261 Pseudomon	356	39	47.6	175	8	ADP46537	Adp46537 Human col
284	40	48.8	264	5	ABW04661	Abw04661 A. thalia	357	39	47.6	184	8	ADW16131	Adw16131 A. thalia
285	40	48.8	286	9	ADW17105	Adw17105 Eucalyptu	358	39	47.6	184	8	ADW02321	Adw02321 Thalecres
286	40	48.8	303	8	ADW87832	Adw87832 Plant ful	359	39	47.6	184	9	AEA26599	Aea26599 Stress to
287	40	48.8	317	9	ADW17108	Adw17108 Eucalyptu	360	39	47.6	194	7	ABW86302	Abw86302 Rice abio
288	40	48.8	328	9	ADW18206	Adw18206 E. grandis	361	39	47.6	196	3	AAW16018	Aaw16018 E. coli p
289	40	48.8	366	6	ABW55427	Abw55427 Amino aci	362	39	47.6	196	4	AAW98320	Aaw98320 Escherich
290	40	48.8	366	10	Aef11588	Aef11588 Tomato me	363	39	47.6	198	7	ABW90275	Abw90275 Rice abio
291	40	48.8	369	8	ADU73566	Adu73566 Hot peppe	364	39	47.6	198	7	ABW90061	Abw90061 Rice abio
292	40	48.8	371	7	ADK72630	Adk72630 AP2 DNA b	365	39	47.6	203	10	Aef28699	Aef28699 Lead Cere
293	40	48.8	372	10	Aef11648	Aef11648 Tomato me	366	39	47.6	203	10	Aef28700	Aef28700 Lead Cere
294	40	48.8	379	8	ADW90615	Adw90615 Plant ful	367	39	47.6	205	10	Aef11614	Aef11614 Soybean m
295	40	48.8	493	4	ABW65502	Abw65502 Drosophil	368	39	47.6	207	8	ABO59992	Abw59992 Human gen
296	40	48.8	539	4	ABW04950	Abw04950 Novel hum	369	39	47.6	208	8	ADY05456	Ady05456 Plant ful
297	40	48.8	582	4	ABW04945	Abw04945 Novel hum	370	39	47.6	208	10	Aef29342	Aef29342 Lead Cere
298	40	48.8	582	7	ABW88330	Abw88330 Rice abio	371	39	47.6	211	1	AAP94616	Aap94616 Alkaline
299	40	48.8	584	8	ADY09948	Ady09948 Plant ful	372	39	47.6	225	6	AAW38264	Aaw38264 Rice dise
300	40	48.8	599	4	ABG18132	Abg18132 Novel hum	373	39	47.6	225	8	ADW68974	Adw68974 Plant ful
301	40	48.8	636	8	ADW73228	Adw73228 Plant ful	374	39	47.6	231	9	AEA34962	Aea34962 Nicotiana
302	40	48.8	723	9	ABW94228	Abw94228 M. xanthu	375	39	47.6	240	10	Aef28701	Aef28701 Lead Cere
303	40	48.8	773	9	AEC08348	Aec08348 A. Evansi	376	39	47.6	240	10	Aef28165	Aef28165 Lead Cere
304	40	48.8	797	4	ABG18136	Abg18136 Novel hum	377	39	47.6	241	10	Aef11546	Aef11546 Rape meth
305	40	48.8	823	4	ABG09263	Abg09263 Novel hum	378	39	47.6	244	10	Aef11532	Aef11532 Rape meth
306	40	48.8	1128	2	AAW05107	Aaw05107 Sequence	379	39	47.6	244	10	Aef11538	Aef11538 Rape meth
307	40	48.8	1666	7	ADG71666	Adg71666 Chlamydom	380	39	47.6	245	10	Aef28164	Aef28164 Lead Cere
308	40	48.8	1779	8	ADH56614	Adh56614 Deduced p	381	39	47.6	262	8	ADW73070	Adw73070 Plant ful
309	40	48.8	1856	3	AAW21802	Aaw21802 B. subtil	382	39	47.6	265	2	AAW10011	Aaw10011 Type 4 GX
310	40	48.8	1856	3	AAW83270	Aaw83270 Polypepti	383	39	47.6	265	2	AAW53294	Aaw53294 Igt-blindi
311	40	48.8	2152	2	AAW39298	Aaw39298 SpnB a po	384	39	47.6	273	8	ADW02598	Adw02598 Thalecres
312	40	48.8	2152	4	AAW70966	Adw70966 S. spinos	385	39	47.6	273	8	ADW62233	Adw62233 Transcript
313	40	48.8	3295	8	ADH39702	Adh39702 Streptomy	386	39	47.6	279	6	ABU40429	Abu40429 Protein e
314	40	48.8	3651	8	ADH39704	Adh39704 Streptomy	387	39	47.6	279	6	ADH86062	Adh86062 Enterococ
315	40	48.8	4924	4	AAW70968	Aaw70968 S. spinos	388	39	47.6	280	6	ABU41977	Abu41977 Protein e

389	39	47.6	315	9	ADM17543	Adw17543 Pinus rad	462	39	47.6	382	2	AAR74223	Aar74223 B. amy101
390	39	47.6	327	10	AEF11566	Aef11566 Tomato me	463	39	47.6	382	2	AAR00247	Aar00247 Subtilisi
391	39	47.6	332	9	ABM93316	Abm93316 M. xanthu	464	39	47.6	382	2	AAM00246	Aam00246 Subtilisi
392	39	47.6	342	4	AAU49617	Aau49617 Propionib	465	39	47.6	382	2	AAR86876	Aar86876 Subtilisi
393	39	47.6	342	6	ABM46136	Abm46136 Propionib	466	39	47.6	382	2	AAR96237	Aar96237 Wild type
394	39	47.6	359	8	ADM19790	Adm19790 Pinus rad	467	39	47.6	382	2	AAR34889	Aar34889 Bacillus
395	39	47.6	366	8	ADM18776	Adm18776 Bacterial	468	39	47.6	382	2	AAR34776	Aar34776 Bacillus
396	39	47.6	372	8	ADP84789	Adp84789 Glycoside	469	39	47.6	382	2	AAY25154	Aay25154 Bacillus
397	39	47.6	373	2	AAR86878	Aar86878 Subtilisi	470	39	47.6	382	2	AAY39228	Aay39228 Bacillus
398	39	47.6	373	2	AAR86877	Aar86877 Subtilisi	471	39	47.6	382	2	AAY16769	Aay16769 B. amy101
399	39	47.6	373	2	AAR86879	Aar86879 Subtilisi	472	39	47.6	382	2	AAY17097	Aay17097 B. amy101
400	39	47.6	373	2	AAR96240	Aar96240 Mutant su	473	39	47.6	382	2	AAY08309	Aay08309 B. amy101
401	39	47.6	373	2	AAR96244	Aar96244 Mutant su	474	39	47.6	382	2	AAR92325	Aar92325 Subtilisi
402	39	47.6	373	2	AAR96245	Aar96245 Mutant su	475	39	47.6	382	2	AAR82793	Aar82793 B. amy101
403	39	47.6	373	2	AAR96242	Aar96242 Mutant su	476	39	47.6	382	2	AAY08314	Aay08314 B. amy101
404	39	47.6	373	2	AAR96241	Aar96241 Mutant su	477	39	47.6	382	2	AAY16767	Aay16767 B. amy101
405	39	47.6	373	2	AAR96241	Aar96241 Mutant su	478	39	47.6	382	3	AAB02972	Aab02972 Bacillus
406	39	47.6	373	2	AAR96243	Aar96243 Mutant su	479	39	47.6	382	3	AAY54618	Aay54618 B. amy101
407	39	47.6	373	2	AAR96239	Aar96239 Mutant su	480	39	47.6	382	3	AAY66980	Aay66980 Bacillus
408	39	47.6	373	2	AAW41692	Aaw41692 BPN' subs	481	39	47.6	382	3	AAY77001	Aay77001 Bacillus
409	39	47.6	373	5	ABB79110	Abb79110 Mutant su	482	39	47.6	382	3	AAB03772	Aab03772 Subtilisi
410	39	47.6	373	5	ABB79092	Abb79092 Mutant su	483	39	47.6	382	5	ABB75071	Abb75071 Bacillus
411	39	47.6	373	5	ABB79094	Abb79094 Mutant su	484	39	47.6	382	5	ABB75072	Abb75072 Bacillus
412	39	47.6	373	5	ABB79101	Abb79101 Mutant su	485	39	47.6	382	5	AAE22942	Aae22942 Bacillus
413	39	47.6	373	5	ABB79107	Abb79107 Mutant su	486	39	47.6	382	5	ABB79085	Abb79085 Bacillus
414	39	47.6	373	5	ABB79100	Abb79100 Mutant su	487	39	47.6	382	5	ABB79086	Abb79086 Mutant su
415	39	47.6	373	5	ABB79108	Abb79108 Mutant su	488	39	47.6	382	5	ABB79088	Abb79088 Mutant su
416	39	47.6	373	5	ABB79109	Abb79109 Mutant su	489	39	47.6	382	5	ABB79087	Abb79087 Mutant su
417	39	47.6	373	5	ABB79093	Abb79093 Mutant su	490	39	47.6	382	6	ABP54389	Abp54389 Bacillus
418	39	47.6	373	5	ABB79117	Abb79117 Mutant su	491	39	47.6	382	6	ADA19154	Ada19154 B. amy101
419	39	47.6	373	5	ABB79123	Abb79123 Mutant su	492	39	47.6	382	6	ADA50344	Ada50344 B. amy101
420	39	47.6	373	5	ABB79097	Abb79097 Mutant su	493	39	47.6	382	6	ADA50348	Ada50348 Bacillus
421	39	47.6	373	5	ABB79120	Abb79120 Mutant su	494	39	47.6	382	6	ADA50346	Ada50346 B. amy101
422	39	47.6	373	5	ABB79091	Abb79091 Mutant su	495	39	47.6	382	6	ADA50345	Ada50345 B. amy101
423	39	47.6	373	5	ABB79102	Abb79102 Mutant su	496	39	47.6	382	7	ADB61326	Adb61326 Protein o
424	39	47.6	373	5	ABB79115	Abb79115 Mutant su	497	39	47.6	382	7	ADB99261	Adb99261 Bacillus
425	39	47.6	373	5	ABB79090	Abb79090 Mutant su	498	39	47.6	382	7	ADC51778	Adc51778 B. amy101
426	39	47.6	373	5	ABB79113	Abb79113 Mutant su	499	39	47.6	382	7	ADE25789	Ade25789 B. amy101
427	39	47.6	373	5	ABB79121	Abb79121 Mutant su	500	39	47.6	382	8	ADH39839	Adh39839 B. lichen
428	39	47.6	373	5	ABB79099	Abb79099 Mutant su	501	39	47.6	382	8	ADH69155	Adh69155 B. amy101
429	39	47.6	373	5	ABB79111	Abb79111 Mutant su	502	39	47.6	382	8	ADJ77871	Adj77871 B. amy101
430	39	47.6	373	5	ABB79122	Abb79122 Mutant su	503	39	47.6	382	8	ADJ46879	Adj46879 B. amy101
431	39	47.6	373	5	ABB79125	Abb79125 Mutant su	504	39	47.6	382	8	ADK40872	Adk40872 B. subtil
432	39	47.6	373	5	ABB79095	Abb79095 Mutant su	505	39	47.6	382	8	ADO04599	Ado04599 Bacillus
433	39	47.6	373	5	ABB79103	Abb79103 Mutant su	506	39	47.6	382	8	ABG91092	Abg91092 Bacillus
434	39	47.6	373	5	ABB79104	Abb79104 Mutant su	507	39	47.6	387	5	ADK68874	Adk68874 Plant ful
435	39	47.6	373	5	ABB79119	Abb79119 Mutant su	508	39	47.6	392	7	ADD84887	Add84887 Bacillus
436	39	47.6	373	5	ABB79124	Abb79124 Mutant su	509	39	47.6	393	4	AUA38518	Aua38518 B. amy101
437	39	47.6	373	5	ABB79098	Abb79098 Mutant su	510	39	47.6	399	3	AAB07808	Aab07808 A. galacta
438	39	47.6	373	5	ABB79114	Abb79114 Mutant su	511	39	47.6	399	8	ADP84787	Adp84787 Bacillus
439	39	47.6	373	5	ABB79116	Abb79116 Mutant su	512	39	47.6	409	8	ADK69035	Adk69035 Plant ful
440	39	47.6	373	5	ABB79096	Abb79096 Mutant su	513	39	47.6	410	6	ABU35825	Abu35825 Protein e
441	39	47.6	373	5	ABB79112	Abb79112 Mutant su	514	39	47.6	437	8	ADY08369	Ady08369 Plant ful
442	39	47.6	373	5	ABB79089	Abb79089 Mutant su	515	39	47.6	448	9	ADW17549	Adw17549 Pinus rad
443	39	47.6	373	5	ABB79105	Abb79105 Mutant su	516	39	47.6	501	3	AAB32978	Aab32978 Pinus rad
444	39	47.6	373	5	ABB79118	Abb79118 Mutant su	517	39	47.6	530	9	ADW18361	Adw18361 Pinus rad
445	39	47.6	379	4	AAAG81170	Aag81170 Mycobacte	518	39	47.6	630	5	AAE18651	Aae18651 Human G-p
446	39	47.6	379	6	ABU36724	Abu36724 Protein e	519	39	47.6	650	7	ADD46694	Add46694 Human Pro
447	39	47.6	379	6	ABU34410	Abu34410 Protein e	520	39	47.6	650	7	ADD46698	Add46698 Human Pro
448	39	47.6	380	6	AAE29952	Aae29952 Bacillus	521	39	47.6	739	6	AAR41080	Aar41080 Human MAP
449	39	47.6	381	2	AAR34464	Aar34464 Bacillus	522	39	47.6	744	4	AAM78406	Aam78406 Human pro
450	39	47.6	382	1	AAP50880	Aap50880 Sequence	523	39	47.6	744	6	ABR41079	Abr41079 Human MAP
451	39	47.6	382	1	AAP70486	Aap70486 Bacillus	524	39	47.6	744	6	ABR41081	Abr41081 Mouse MAP
452	39	47.6	382	1	AAP70052	Aap70052 Bacillus	525	39	47.6	744	6	ABR41078	Abr41078 Human MAP
453	39	47.6	382	1	AAP71731	Aap71731 Subtilisi	526	39	47.6	744	7	AAE38299	Aae38299 Human ves
454	39	47.6	382	1	AAP71732	Aap71732 Subtilisi	527	39	47.6	744	7	ADD46692	Add46692 Rat Prote
455	39	47.6	382	1	AAP71733	Aap71733 Encodes s	528	39	47.6	744	7	ADD46696	Add46696 Rat Prote
456	39	47.6	382	1	AAP80271	Aap80271 Amino aci	529	39	47.6	744	7	ADE55423	Ade55423 Human Pro
457	39	47.6	382	2	AAR03736	Aar03736 Subtilisi	530	39	47.6	744	7	ADE55426	Ade55426 Human Pro
458	39	47.6	382	2	AAR58691	Aar58691 Subtilisi	531	39	47.6	744	7	ADJ69221	Adj69221 Human hea
459	39	47.6	382	2	AAR70012	Aar70012 Subtilisi	532	39	47.6	744	8	ADR89490	Adr89490 Apoptosis
460	39	47.6	382	2	AAR75161	Aar75161 B. amy101	533	39	47.6	744	8	ADR89492	Adr89492 Apoptosis
461	39	47.6	382	2	AAR65247	Aar65247 B. amy101	534	39	47.6	744	9	ADX26340	Adx26340 Novel cel

535	39	47.6	744	9	ADX26414	Novel cel	608	38	46.3	189	8	ADQ65251	Novel hum
536	39	47.6	744	9	ADX26271	Novel cel	609	38	46.3	192	4	ABA48292	Human ZF4
537	39	47.6	751	4	ABG04583	Novel hum	610	38	46.3	193	3	AGS60567	Arabidops
538	39	47.6	751	4	ABG16291	Novel hum	611	38	46.3	197	8	ADX68833	Plant ful
539	39	47.6	762	4	AAW79390	Human pro	612	38	46.3	201	9	ADM18204	E. grandis
540	39	47.6	764	8	ADR66089	Human pro	613	38	46.3	202	8	ADT60371	Plant pol
541	39	47.6	764	8	ADR66431	Human pro	614	38	46.3	210	2	AAW31709	X25 gene
542	39	47.6	846	8	ADO61707	Transcrip	615	38	46.3	210	3	AAV97227	Plant tra
543	39	47.6	865	9	AED07619	Aerononas	616	38	46.3	210	7	ADJ69448	Human hea
544	39	47.6	866	2	AAW02159	Soluble c	617	38	46.3	216	8	ADT58219	Plant pol
545	39	47.6	866	3	AAV52307	Vibrio fu	618	38	46.3	216	8	ADG69005	Plant ful
546	39	47.6	879	6	ABJ25894	Aspergill	619	38	46.3	218	3	AGS50951	Arabidops
547	39	47.6	988	4	ABG16888	Novel hum	620	38	46.3	218	3	AGS09633	Arabidops
548	39	47.6	988	7	ADI21719	Novel hum	621	38	46.3	218	4	AAE01899	Arabidops
549	39	47.6	1032	6	ABJ26494	Aspergill	622	38	46.3	218	5	AAU93058	Arabidops
550	39	47.6	1366	4	AAE02536	Forcine a	623	38	46.3	218	5	ADD30029	Plant yie
551	39	47.6	1372	7	ADT59683	Rat Prote	624	38	46.3	218	7	ADG37271	Plant yie
552	39	47.6	1372	7	ADD45148	Rat Prote	625	38	46.3	218	7	ADE31503	Plant yie
553	39	47.6	1372	7	ADD45604	Rat Prote	626	38	46.3	218	8	ADI41789	Plant tra
554	39	47.6	1372	7	ADT59687	Rat Prote	627	38	46.3	218	8	ADO02387	Thalecres
555	39	47.6	1372	7	ADD47529	Rat Prote	628	38	46.3	218	9	ADV42286	Plant tra
556	39	47.6	1373	5	ABB57364	Mouse isc	629	38	46.3	218	9	ADM73200	Novel pla
557	39	47.6	1373	9	ADW44460	Murine pr	630	38	46.3	218	9	AEA26691	Stress to
558	39	47.6	1373	10	AEF119250	Mus muscu	631	38	46.3	218	9	AEA27259	Stress to
559	38.5	47.0	142	2	AAE50193	Heavy cha	632	38	46.3	218	9	AED44057	Floral-bu
560	38.5	47.0	142	2	AAV26585	Heavy cha	633	38	46.3	218	10	AEF28163	Lead_Cere
561	38.5	47.0	142	8	ADH87532	Humanised	634	38	46.3	218	10	AEF28705	Lead_Cere
562	38.5	47.0	734	9	ABM93705	M. xanthu	635	38	46.3	218	10	AEF28704	Lead_Cere
563	38	46.3	9	3	AAV70256	Peptide e	636	38	46.3	227	9	ABM96418	M. xanthu
564	38	46.3	33	10	AEF11616	Human PY	637	38	46.3	229	10	AEF11616	Soybean m
565	38	46.3	35	10	AEF28703	Human PY	638	38	46.3	229	10	AEF28703	Lead_Cere
566	38	46.3	53	10	AEF28703	Human PY	639	38	46.3	232	3	AGS09632	Arabidops
567	38	46.3	62	4	ADH75729	Thalecres	640	38	46.3	232	3	AGS09505	Arabidops
568	38	46.3	62	8	ADH75729	Thalecres	641	38	46.3	238	7	ABO75119	Pseudomon
569	38	46.3	68	2	AAV31475	A. thalia	642	38	46.3	246	8	ADM72393	A. thalia
570	38	46.3	68	2	AAV31475	A. thalia	643	38	46.3	246	2	AAW37091	Lycopersi
571	38	46.3	68	3	ABZ25834	AP2 domai	644	38	46.3	248	3	AGS07101	Arabidops
572	38	46.3	68	3	ABZ25831	AP2 domai	645	38	46.3	248	3	AGS09632	Arabidops
573	38	46.3	74	5	ABP01080	Human ORF	646	38	46.3	248	4	AAE02504	Arabidops
574	38	46.3	90	7	ABO79997	Pseudomon	647	38	46.3	248	7	ABO43103	A. thalia
575	38	46.3	102	5	AAO17456	Human liv	648	38	46.3	248	7	ADE37273	Plant yie
576	38	46.3	124	8	ADI43274	Plant tra	649	38	46.3	248	8	ADI41541	Plant tra
577	38	46.3	124	8	ADO63662	Transcrip	650	38	46.3	248	8	ADO01609	Thalecres
578	38	46.3	124	9	AEA26233	Stress to	651	38	46.3	248	8	ADM72394	A. thalia
579	38	46.3	124	9	AEA27245	Stress to	652	38	46.3	248	8	ADN72811	Thale cre
580	38	46.3	124	10	AEF28687	Lead_Cere	653	38	46.3	248	9	ADV42282	Plant tra
581	38	46.3	128	10	AEF28684	Lead_Cere	654	38	46.3	249	3	AAG45281	Arabidops
582	38	46.3	138	8	ADH09769	Human hos	655	38	46.3	254	8	ADY04736	Plant ful
583	38	46.3	138	8	ADO63752	Transcrip	656	38	46.3	254	8	ADX68834	Plant ful
584	38	46.3	138	9	AEA26245	Stress to	657	38	46.3	256	8	ADX71795	Plant ful
585	38	46.3	138	9	AEA27244	Stress to	658	38	46.3	260	8	ABX56932	Human pro
586	38	46.3	138	10	AEF28681	Lead_Cere	659	38	46.3	260	8	ADX68572	Plant ful
587	38	46.3	139	7	ADD30332	Plant yie	660	38	46.3	260	10	AEF11640	Tomato me
588	38	46.3	139	8	ADI43671	Plant tra	661	38	46.3	266	4	AAU43199	Propionib
589	38	46.3	139	8	ADO62995	Transcrip	662	38	46.3	266	6	ABM39718	Propionib
590	38	46.3	139	9	AEA27235	Stress to	663	38	46.3	266	10	AEF11536	Rape meth
591	38	46.3	139	9	AEA26193	Stress to	664	38	46.3	268	3	AGS07100	Arabidops
592	38	46.3	139	10	AEF28685	Lead_Cere	665	38	46.3	268	3	AGS49621	Arabidops
593	38	46.3	141	3	AGS45282	Arabidops	666	38	46.3	268	4	ABA47310	CBP-1 8/
594	38	46.3	141	4	AAE03578	Human pro	667	38	46.3	279	8	ADX65866	Plant ful
595	38	46.3	141	8	ADO63742	Transcrip	668	38	46.3	281	8	ADT56727	Plant pol
596	38	46.3	141	9	ADV42287	Plant tra	669	38	46.3	282	3	AGS45280	Arabidops
597	38	46.3	141	9	AEA27247	Stress to	670	38	46.3	282	8	ADO03305	Thalecres
598	38	46.3	141	9	AEA26235	Stress to	671	38	46.3	282	8	ADO62885	Transcrip
599	38	46.3	143	4	AAO13070	Human pol	672	38	46.3	282	9	ADV42288	Plant tra
600	38	46.3	144	6	ADA54580	Human pro	673	38	46.3	285	10	AEF11544	Soybean m
601	38	46.3	148	3	AGS06986	Arabidops	674	38	46.3	290	8	ADX72629	Plant ful
602	38	46.3	155	7	ABO72897	Pseudomon	675	38	46.3	292	4	AAE03573	Human pro
603	38	46.3	159	4	AAE03580	Human pro	676	38	46.3	296	6	ABU97215	Enzyme po
604	38	46.3	172	3	AGS59121	Arabidops	677	38	46.3	300	10	AEF11542	Soybean m
605	38	46.3	172	3	AGS60568	Arabidops	678	38	46.3	305	8	ADX66976	Plant ful
606	38	46.3	173	3	AGS09634	Arabidops	679	38	46.3	308	4	AAW01109	CFE 112 p
607	38	46.3	173	3	AGS50952	Arabidops	680	38	46.3	308	8	ADK46438	Streptoco

681	38	46.3	309	9	ADM17107	Adw17107 Eucalyptu	754	38	46.3	750	7	ADM04450	Adm04450 Human pro
682	38	46.3	310	4	AAE03575	Aae03575 Human pro	755	38	46.3	750	9	AEC87380	Aec87380 Human pro
683	38	46.3	317	4	AAE03579	Aae03579 Human pro	756	38	46.3	750	4	AEC03583	Aae03583 Human pro
684	38	46.3	329	10	AEF11568	Aef11568 Coffee me	757	38	46.3	810	7	ADE62643	Ad62643 Rat Prote
685	38	46.3	331	9	ADR93988	Adr93988 Novel S.	758	38	46.3	810	7	ABE62647	Ab62647 Rat Prote
686	38	46.3	331	9	AEA57858	Aea57858 Streptoco	759	38	46.3	810	8	ADR32025	Adr32025 Rat NELL1
687	38	46.3	333	3	AAE27694	Aae27694 Arabidops	760	38	46.3	852	6	ABU12116	Abu12116 Human pro
688	38	46.3	335	3	AAE03754	Aae03754 Dehydrati	761	38	46.3	879	4	ABG25755	Abg25755 Novel hum
689	38	46.3	335	3	AAE03754	Aae03754 Dehydrati	762	38	46.3	879	4	ABG28383	Abg28383 Novel hum
690	38	46.3	335	3	AAE03754	Aae03754 Dehydrati	763	38	46.3	908	4	AAE03572	Aae03572 Human pro
691	38	46.3	335	3	AAE02539	Aae02539 A. thalia	764	38	46.3	909	8	ADG23793	Adg23793 Bacterial
692	38	46.3	335	7	ADD55736	Add55736 Thalecres	765	38	46.3	950	7	AEE72771	Aee72771 Novel hum
693	38	46.3	335	7	ADD30754	Add30754 Plant yie	766	38	46.3	959	4	ABE6947	Ab6947 Human met
694	38	46.3	335	8	ADI43763	Adi43763 Plant tra	767	38	46.3	971	8	ABM81634	Abm81634 Tumour-as
695	38	46.3	335	8	ADO01629	Ado01629 Thalecres	768	38	46.3	972	6	ABU11811	Abu11811 Human MDD
696	38	46.3	335	8	ADP22718	Adp22718 Arabidops	769	38	46.3	972	6	ABU11669	Abu11669 Human MDD
697	38	46.3	335	8	ADG73829	Adg73829 A. thalia	770	38	46.3	1010	8	ADO05302	Ado05302 Chlamydom
698	38	46.3	335	8	ADS19137	Ads19137 Rice DREB	771	38	46.3	1021	8	ADO07041	Ado07041 Human pro
699	38	46.3	335	8	ADT06981	Adt06981 Arabidops	772	38	46.3	1098	7	ADB64321	Adb64321 Human pro
700	38	46.3	335	9	AEA26985	Aea26985 Stress to	773	38	46.3	1117	7	ADZ44782	Adz44782 Human ang
701	38	46.3	335	10	AEF31036	Aef31036 Arabidops	774	38	46.3	1120	5	ABG76502	Abg76502 DNA encod
702	38	46.3	342	7	ADJ11386	Adj11386 Rice prot	775	38	46.3	1124	8	ADT89120	Adt89120 Tomato MS
703	38	46.3	342	7	ADJ11726	Adj11726 Rice prot	776	38	46.3	1212	6	ABU40820	Abu40820 Protein e
704	38	46.3	342	7	ABM85988	Abm85988 Rice abio	777	38	46.3	1213	7	ADF06193	Adf06193 Bacterial
705	38	46.3	342	7	ABM90033	Abm90033 Rice abio	778	38	46.3	1343	8	ADS23976	Ads23976 Bacterial
706	38	46.3	342	10	AEF11628	Aef11628 Rice meth	779	38	46.3	1404	4	ABG14929	Abg14929 Novel hum
707	38	46.3	346	8	ADY06604	Ady06604 Plant ful	780	38	46.3	1558	8	ADT89145	Adt89145 MSH1conse
708	38	46.3	346	8	ADY24678	Ady24678 Plant ful	781	37.5	45.7	124	4	AUS58334	Aus58334 Propionib
709	38	46.3	355	8	ADX68445	Adx68445 Plant ful	782	37.5	45.7	124	6	ABM54853	Abm54853 Propionib
710	38	46.3	356	4	AAE03581	Aae03581 Human pro	783	37.5	45.7	152	6	AAO30394	Aao30394 Rabbit 14
711	38	46.3	358	3	AAE03581	Aae03581 Arabidops	784	37.5	45.7	152	8	ADO79140	Ado79140 Anti-CD83
712	38	46.3	358	3	AAE03581	Aae03581 Arabidops	785	37.5	45.7	290	3	AY95037	Ay95037 Candida a
713	38	46.3	358	7	ADB31779	Adb31779 Plant (A.	786	37.5	45.7	291	10	AEF79950	Aef79950 Growth-/y
714	38	46.3	358	7	ADY55714	Ady55714 Thalecres	787	37.5	45.7	454	6	AAO30347	Aao30347 Rabbit 14
715	38	46.3	358	8	ADO01607	Ado01607 Thalecres	788	37.5	45.7	454	8	ADO79097	Ado79097 Rabbit an
716	38	46.3	358	10	AEF11592	Aef11592 A. thalia	789	37.5	45.7	1401	2	AY14519	Ay14519 Mouse WRN
717	38	46.3	358	10	AEF11644	Aef11644 A. thalia	790	37	45.1	24	4	ABM85520	Abm85520 Heparanas
718	38	46.3	362	8	ADX93368	Adx93368 Plant ful	791	37	45.1	56	8	ABO54957	Abos4957 Human gen
719	38	46.3	362	10	AEF11666	Aef11666 Rice meth	792	37	45.1	65	9	ADV42293	Adv42293 Plant AP2
720	38	46.3	364	8	ADY05458	Ady05458 Plant ful	793	37	45.1	65	9	ADV42315	Adv42315 Plant AP2
721	38	46.3	364	8	ADX95844	Adx95844 Plant ful	794	37	45.1	65	9	ADV42291	Adv42291 Plant AP2
722	38	46.3	364	8	ADX95845	Adx95845 Plant ful	795	37	45.1	65	9	ADV42296	Adv42296 Plant AP2
723	38	46.3	375	10	AEF11654	Aef11654 Beech met	796	37	45.1	65	9	ADV42292	Adv42292 Plant AP2
724	38	46.3	378	8	ADX93676	Adx93676 Plant ful	797	37	45.1	65	9	ADV42294	Adv42294 Plant AP2
725	38	46.3	378	10	AEF11590	Aef11590 Beech met	798	37	45.1	65	9	ADV42313	Adv42313 Plant AP2
726	38	46.3	379	8	ADS42376	Ads42376 Bacterial	799	37	45.1	66	4	AAU51977	Aau51977 Propionib
727	38	46.3	387	3	AB32761	Ab32761 Eucalyptu	800	37	45.1	66	6	ABM48496	Abm48496 Propionib
728	38	46.3	387	10	AEF11634	Aef11634 Tobacco m	801	37	45.1	68	4	ABM59552	Abm59552 Propionib
729	38	46.3	389	8	ADX93781	Adx93781 Plant ful	802	37	45.1	68	6	ABM59552	Abm59552 Propionib
730	38	46.3	397	6	AAE29815	Aae29815 Streptoco	803	37	45.1	72	5	ABP00912	Abp00912 Human ORF
731	38	46.3	418	8	ADY09711	Ady09711 Plant ful	804	37	45.1	73	3	ABM28578	Abm28578 Human SCA
732	38	46.3	421	8	ADY24686	Ady24686 Plant ful	805	37	45.1	75	3	AAO00782	Aao00782 Human sec
733	38	46.3	422	8	ADY05453	Ady05453 Plant ful	806	37	45.1	76	8	ADM96968	Adm96968 Human pan
734	38	46.3	438	4	AAE03582	Aae03582 Human pro	807	37	45.1	82	4	AAU92673	Aau92673 Human dig
735	38	46.3	462	7	ADG63510	Adg63510 Rat Prote	808	37	45.1	84	5	ABP31965	Abp31965 Human ORF
736	38	46.3	468	4	AAE03574	Aae03574 Human pro	809	37	45.1	100	8	ADY22704	Ady22704 Plant ful
737	38	46.3	468	7	ABO83691	Abos83691 Pseudomon	810	37	45.1	125	5	ADK34873	Adk34873 Novel hum
738	38	46.3	470	2	AAW55998	Aaw55998 Protein S	811	37	45.1	127	9	AEA27236	Aea27236 Stress to
739	38	46.3	470	9	ADY27582	Ady27582 Neuramini	812	37	45.1	131	5	AAU93095	Aau93095 Arabidops
740	38	46.3	470	9	ADY27580	Ady27580 Neuramini	813	37	45.1	131	7	ADD30336	Add30336 Plant yie
741	38	46.3	470	9	ADY27578	Ady27578 Neuramini	814	37	45.1	131	8	ADI43673	Adi43673 Plant tra
742	38	46.3	471	2	ADH29824	Adh29824 Swinepox	815	37	45.1	131	8	ADI43273	Adi43273 Plant tra
743	38	46.3	493	9	ADW17552	Adw17552 Pinus rad	816	37	45.1	131	8	ADO62999	Ado62999 Transcript
744	38	46.3	506	7	ABO78690	Abos78690 Pseudomon	817	37	45.1	131	8	ADO63660	Ado63660 Transcript
745	38	46.3	507	4	AAE03576	Aae03576 Human pro	818	37	45.1	131	9	AEA26195	Aea26195 Stress to
746	38	46.3	531	8	ADO13858	Ado13858 Protein e	819	37	45.1	131	9	AEA26231	Aea26231 Stress to
747	38	46.3	570	3	AAO30056	Aao30056 Arabidops	820	37	45.1	131	9	AEA27241	Aea27241 Stress to
748	38	46.3	589	4	AAE03577	Aae03577 Human pro	821	37	45.1	131	10	AEF28691	Aef28691 Lead Cere
749	38	46.3	603	3	AAE03055	Aae03055 Arabidops	822	37	45.1	131	10	AEF28686	Aef28686 Lead Cere
750	38	46.3	613	4	ABE62232	Ab62232 Drosophyl	823	37	45.1	131	10	AEF28683	Aef28683 Lead Cere
751	38	46.3	618	3	AAO30054	Aao30054 Arabidops	824	37	45.1	133	9	AEA27237	Aea27237 Stress to
752	38	46.3	662	2	AAU14585	Aau14585 A.thalian	825	37	45.1	137	6	ABM70003	Abm70003 Photornab
753	38	46.3	662	5	ABB93201	Abb93201 Herbicida	826	37	45.1	140	7	ABO61374	Abos61374 Klebsiell

827	37	45.1	141	9	ABM94687	Am94687 M. xanthu	900	37	45.1	450	7	ADG10792	AdG10792 Human STA
828	37	45.1	148	4	AAO06361	Aao06361 Human pol	901	37	45.1	450	7	ADG10790	AdG10790 Human STA
829	37	45.1	149	9	AEA26319	Aea26319 Stress to	902	37	45.1	451	6	ABU89756	Abu89756 Protein d
831	37	45.1	149	9	AEA27252	Aea27252 Stress to	903	37	45.1	451	6	ABU19526	Abu19526 Protein e
832	37	45.1	151	8	ADO63750	Ado63750 Transcrip	904	37	45.1	463	8	ADX74862	Adx74862 Plant ful
833	37	45.1	151	9	AEA26243	Aea26243 Stress to	905	37	45.1	466	7	ADG33854	AdG33854 Actinomyc
834	37	45.1	151	9	AEA27251	Aea27251 Stress to	906	37	45.1	480	8	ADT58933	Adt58933 Plant pol
835	37	45.1	158	8	ADO63748	Ado63748 Transcrip	907	37	45.1	487	4	AAV72635	Aay72635 Exophiala
836	37	45.1	158	9	AEA27250	Aea27250 Stress to	908	37	45.1	487	5	AAE26385	Aae26385 Exophiala
837	37	45.1	158	9	AEA26241	Aea26241 Stress to	909	37	45.1	487	5	ABG75928	Abg75928 Fumoniain
838	37	45.1	160	4	AAU56800	Aau56800 Propionib	910	37	45.1	487	6	ABU07912	Abu07912 Exophiala
839	37	45.1	161	2	AAW37090	Aaw37090 Lycopersi	911	37	45.1	487	6	ABU62939	Abu62939 E. spinif
840	37	45.1	162	6	ABO14768	Abol14768 Novel hum	912	37	45.1	494	8	ADS28840	Ads28840 Bacterial
841	37	45.1	180	8	ADX91956	Adx91956 Plant ful	913	37	45.1	538	7	ADE08404	Ade08404 Novel pro
842	37	45.1	181	9	ADW17088	Adw17088 Eucalyptu	914	37	45.1	545	9	AEA21110	Aea21110 Novel hum
843	37	45.1	191	2	AAR39562	Aar39562 Sequence	915	37	45.1	574	9	ABM72146	Abm72146 Streptomy
844	37	45.1	196	6	ABR53581	Abr53581 Protein s	916	37	45.1	616	6	ABM72146	Abm72146 Streptomy
845	37	45.1	196	7	ADK64334	Adk64334 Disease t	917	37	45.1	649	8	ADN23351	Adn23351 Bacterial
846	37	45.1	196	9	AEA62715	Aea62715 Mitochond	918	37	45.1	663	8	ADS44672	Ads44672 Bacterial
847	37	45.1	222	8	ADX68377	Adx68377 Plant ful	919	37	45.1	686	7	ADC95375	Adc95375 E. faeciu
848	37	45.1	226	9	AED26774	Aed26774 DNA-PKcs	920	37	45.1	703	8	ADS15435	Ads15435 HSV-1 pol
849	37	45.1	233	6	ADA36235	Ada36235 Acinetoba	921	37	45.1	735	2	AAW69761	Aaw69761 Acetobact
850	37	45.1	233	6	ADG03716	Adg03716 Novel hum	922	37	45.1	739	4	ADR86165	Adr86165 Aspergill
851	37	45.1	237	8	ADL42039	Adl42039 Plant tra	923	37	45.1	781	4	ABR58388	Abbr58388 Drosophil
852	37	45.1	237	8	ADL42039	Adl42039 Plant tra	924	37	45.1	823	5	ABP65373	Abp65373 Bifidbac
853	37	45.1	243	7	ADW17540	Adw17540 Pinus rad	925	37	45.1	903	6	ABU38811	Abu38811 Protein e
854	37	45.1	244	9	ADV42266	Adv42266 Plant tra	926	37	45.1	974	5	ABP35581	Abp35581 Fungal 2B
855	37	45.1	247	9	ADV42266	Adv42266 Plant tra	927	37	45.1	988	7	ABO78500	Abp78500 Pseudomon
856	37	45.1	256	3	AGL13479	Agil13479 Arabidops	928	37	45.1	1005	7	ADE98304	Ade98304 Cancer-li
857	37	45.1	259	8	ADL42038	Adl42038 Plant tra	929	37	45.1	1032	7	ADE98279	Ade98279 Cancer-li
858	37	45.1	259	8	ADL42038	Adl42038 Plant tra	930	37	45.1	1039	7	ADE98283	Ade98283 Cancer-li
859	37	45.1	265	7	ABM74280	Abm74280 DNA clone	931	37	45.1	1040	7	ADE98289	Ade98289 Cancer-li
860	37	45.1	267	8	ADX90931	Adx90931 Plant ful	932	37	45.1	1063	7	ADE98280	Ade98280 Cancer-li
861	37	45.1	270	5	ABR53670	Abbr53670 Lactococc	933	37	45.1	1066	7	ADE98302	Ade98302 Cancer-li
862	37	45.1	270	9	ADV42236	Adv42236 Plant tra	934	37	45.1	1089	7	ADE98305	Ade98305 Cancer-li
863	37	45.1	271	8	ADM48058	Adm48058 Polypepti	935	37	45.1	1125	7	ADE98293	Ade98293 Cancer-li
864	37	45.1	271	9	ADV42254	Adv42254 Plant tra	936	37	45.1	1152	7	ADE98288	Ade98288 Cancer-li
865	37	45.1	272	9	ADV42256	Adv42256 Plant tra	937	37	45.1	1159	7	ADE98281	Ade98281 Cancer-li
866	37	45.1	274	9	ADW17540	Adw17540 Pinus rad	938	37	45.1	1162	8	ADL16264	Adl16264 Human nuc
867	37	45.1	276	8	ADK66569	Adk66569 Plant ful	939	37	45.1	1178	6	ABU32204	Abu32204 Protein e
868	37	45.1	278	10	AEF11626	Aef11626 Soybean m	940	37	45.1	1178	7	ABO63427	Abob63427 Klebsiell
869	37	45.1	278	8	ADK71478	Adk71478 Plant ful	941	37	45.1	1186	6	ABR41085	Abrr41085 Human MAP
870	37	45.1	288	4	ABG09003	Abg09003 Novel hum	942	37	45.1	1186	6	ABR41085	Abrr41085 Human MAP
871	37	45.1	288	4	ABG03400	Abg03400 Novel hum	943	37	45.1	1186	7	ADF09579	Adf09579 Human Sec
872	37	45.1	294	9	ADV42258	Adv42258 Plant tra	944	37	45.1	1186	7	ADE98286	Ade98286 Cancer-li
873	37	45.1	296	5	ADP48345	Adp48345 Staphyloc	945	37	45.1	1186	8	ADI24519	Adi24519 Human mod
874	37	45.1	299	8	ADS21667	Ads21667 Bacterial	946	37	45.1	1186	8	ADU60281	Adu60281 Housekeep
875	37	45.1	303	6	AAE33251	Aae33251 Rice DBP1	947	37	45.1	1186	9	ADX07600	Adx07600 Cyclin-de
876	37	45.1	303	9	ADV42234	Adv42234 Plant tra	948	37	45.1	1186	9	AEEL8644	Aeel8644 Human Pum
877	37	45.1	310	8	ADK95485	Adk95485 Plant ful	949	37	45.1	1186	10	AEF34743	Aef34743 Human PUM
878	37	45.1	311	8	ADK74036	Adk74036 Plant ful	950	37	45.1	1186	10	AEF34745	Aef34745 Human exp
879	37	45.1	323	4	ABB64496	Abb64496 Drosophil	951	37	45.1	1186	10	AEF34747	Aef34747 Human exp
880	37	45.1	327	9	ABE39240	Abe39240 L. pneumo	952	37	45.1	1186	10	AEF34747	Aef34747 Human exp
881	37	45.1	344	9	AEEO2491	Aee02491 Human her	953	37	45.1	1209	9	ADE98292	Ade98292 Human pla
882	37	45.1	344	9	AEEO2491	Aee02491 Human her	954	37	45.1	1441	9	AEW74917	Aew74917 pGJA-P/VA
883	37	45.1	348	9	AEA27258	Aea27258 Stress to	955	37	45.1	1561	2	AAW02098	Aaw02098 S. mutans
884	37	45.1	352	8	ADS15433	Ads15433 HSV-1 pol	956	37	45.1	1561	6	ABU44355	Abu44355 Protein e
885	37	45.1	352	8	ABE35793	Abe35793 L. pneumo	957	37	45.1	1562	9	ABE91501	Abu44355 Microbial
886	37	45.1	353	10	AEF11636	Aef11636 Wheat met	958	37	45.1	1690	4	ABE61144	Abbe61144 Drosophil
887	37	45.1	363	8	ADS41611	Ads41611 Bacterial	959	37	45.1	1690	4	ABE61173	Abbe61173 Drosophil
888	37	45.1	376	8	ADY08195	Ady08195 Plant ful	960	37	45.1	1690	9	ADY85283	Ady85283 Drosophil
889	37	45.1	384	2	ABU21232	Abu21232 Protein e	961	37	45.1	2130	8	ADQ74677	Adq74677 Streptomy
890	37	45.1	386	2	AAU06416	Aau06416 Aspergill	962	37	45.1	3201	4	ADQ74677	Adq74677 Streptomy
891	37	45.1	416	7	ABO80795	Abob80795 Pseudomon	963	37	45.1	3798	3	AAV58577	Aay58577 Soranglum
892	37	45.1	419	8	ADY06605	Ady06605 Plant ful	964	36.5	44.5	76	7	ADJ68338	Adj68338 Human hea
893	37	45.1	442	8	ADS29959	Ads29959 Bacterial	965	36.5	44.5	105	6	ABM53700	Abm53700 Propionib
894	37	45.1	443	8	ADN19253	Adn19253 Bacterial	966	36.5	44.5	105	6	ABM53700	Abm53700 Propionib
895	37	45.1	443	9	ADM43134	Adm43134 Baker's y	967	36.5	44.5	136	6	ABM62364	Abm62364 Propionib
896	37	45.1	447	5	AAE21507	Aae21507 Human gen	968	36.5	44.5	138	4	ABG13594	Abg13594 Novel hum
897	37	45.1	447	5	AAE21507	Aae21507 Human gen	969	36.5	44.5	183	9	ABE36549	Abeg36549 L. pneumo
898	37	45.1	447	8	ADL78238	Adl78238 Albumin f	971	36.5	44.5	183	9	ABE39943	Abeg39943 L. pneumo
899	37	45.1	450	7	ADG10678	Adg10678 Human STA	972	36.5	44.5	188	3	ABA42212	Aab42212 Human ORF

973 36.5 44.5 250 5 ABP45634 Human Bly
 974 36.5 44.5 250 7 ADG96461 Single ch
 975 36.5 44.5 250 9 AED78514 Human B L
 976 36.5 44.5 261 5 ABP43699 Interapti
 977 36.5 44.5 319 6 ABM15877 Mycobacte
 978 36.5 44.5 333 9 AEB49952 E. coli p
 979 36.5 44.5 339 7 ADD30286 Plant yie
 980 36.5 44.5 339 8 ADI44171 Plant tra
 981 36.5 44.5 346 4 AAU57839 Propionib
 982 36.5 44.5 346 6 ABM54358 Propionib
 983 36.5 44.5 362 4 ABG18548 Novel hum
 984 36.5 44.5 363 7 ADM05629 Human pro
 985 36.5 44.5 363 8 ADQ17383 Human sof
 986 36.5 44.5 363 9 AEC88559 Human cDN
 987 36.5 44.5 645 4 ABQ21616 Novel hum
 988 36.5 44.5 909 8 ADT56533 Plant pol
 989 36.5 44.5 9594 10 AEE75860 Streptomy
 990 36 43.9 56 4 AAU43939 Propionib
 991 36 43.9 56 6 ABM40458 Propionib
 992 36 43.9 64 2 AAY59872 Human nor
 993 36 43.9 64 8 ADR90165 Human tum
 994 36 43.9 65 4 AAG75239 Human col
 995 36 43.9 66 2 AAY48484 Human bre
 996 36 43.9 66 4 AAO13377 Human pol
 997 36 43.9 79 4 AAU53913 Propionib
 998 36 43.9 79 6 ABM50432 Propionib
 999 36 43.9 87 4 AAU40108 Propionib
 1000 36 43.9 87 6 ABM36627 Propionib

ALIGNMENTS

RESULT 1
 ADL16423
 ID ADL16423 standard; peptide; 15 AA.
 XX
 AC ADL16423;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Human heparanase antigenic peptide.
 XX
 KW Human; heparanase; heparanase-dependent cancer; cancer;
 KW autoimmune reaction; inflammation; immunogen.
 XX
 OS Homo sapiens.
 XX
 PN US2003236215-A1.
 XX
 PD 25-DEC-2003.
 XX
 PF 09-JUN-2003; 2003US-00456573.
 XX
 PR 31-AUG-1998; 98WO-0017954.
 PR 01-MAR-1999; 99US-00258892.
 PR 08-NOV-1999; 99US-00435739.
 XX
 (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 PI Pecker I, Vlodavsky I, Feinstein E;
 XX
 DR WPI; 2004-070610/07.
 XX
 PT New antisense oligonucleotide hybridizable with a polynucleotide encoding
 PT a polypeptide with heparanase activity, useful for treating diseases such
 PT as cancer and autoimmune disorders.
 XX
 PS Example 17; SEQ ID NO 54; 108pp; English.
 XX
 CC The invention relates to an antisense oligonucleotide (ASO) comprising a
 CC polynucleotide or a polynucleotide analogue of at least 10 bases being

CC hybridisable in vivo , under physiological conditions, with a portion of
 CC a polynucleotide strand encoding a polypeptide having heparanase
 CC catalytic activity. Also included are a method of in vivo downregulating
 CC heparanase activity (comprising administering the ASO in vivo), a method
 CC of treating a subject suffering from a pathological condition
 CC (characterised by heparanase activity, comprising administering ASO to
 CC the subject), a pharmaceutical composition comprising the ASO and a
 CC carrier, an antisense nucleic acid construct (comprising a promoter
 CC sequence and a polynucleotide sequence directing the synthesis of an
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo ,
 CC under physiological conditions, with a polynucleotide strand encoding a
 CC polypeptide having heparanase catalytic activity), a method of in vivo
 CC downregulating heparanase activity (comprising administering in vivo the
 CC antisense nucleic acid construct), a pharmaceutical composition
 CC comprising the antisense nucleic acid construct and a carrier, and an
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
 CC analogue of at least 10 bases being hybridisable in vivo , under
 CC physiological conditions, with a portion of a polynucleotide strand being
 CC characterised by forming at least a portion of an untranslated region
 CC (UTR) for a polynucleotide strand encoding a polypeptide having
 CC heparanase catalytic activity. The methods and compositions of the
 CC present invention are useful for the prevention and/or treatment of
 CC diseases or conditions associated with aberrant heparanase activity, such
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and
 CC inflammation. The present sequence is a human heparanase peptide used
 CC raise anti-heparanase antibodies.
 XX
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 8; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSSAY 15
 DB 1 RPQKKVWLGETSSAY 15
 RESULT 2
 ADR88215
 ID ADR88215 standard; peptide; 15 AA.
 XX
 AC ADR88215;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human heparanase epitope pep9.
 XX
 KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
 KW haematopoietic malignancy; lymphocytic leukemia; myelogenous leukaemia;
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
 KW human; heparanase; enzyme; epitope.
 XX
 OS Homo sapiens.
 XX
 PN US2004170631-A1.
 XX
 PD 02-SEP-2004.
 XX
 PF 28-NOV-2003; 2003US-00722502.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 PR 22-AUG-2003; 2003US-00645659.
 XX
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VSEL/) VAN GELDER J. M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
DR WPI; 2004-625084/60.
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX
PS Claim 7; SEQ ID NO 9; 58pp; English.
XX
CC The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an antibody or its portion
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease), for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is human heparanase epitope.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 82; DB 8; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RPQKKVWLGETSSAY 15
| | | | | | | | | | | | | | |
Db 1 RPQKKVWLGETSSAY 15
| | | | | | | | | | | | | | |
RESULT 3
ADT78182
ID ADT78182 standard; peptide; 15 AA.
XX
AC ADT78182;
XX
XX 13-JAN-2005 (first entry)
XX
DE Functional peptide epitope of human heparanase, pep9.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
XX
OS Homo sapiens.
XX
XX US2004213789-A1.
XX
XX 28-OCT-2004.
XX
XX 22-AUG-2003; 2003US-00645659.
XX
XX 02-SEP-1997; 97US-00922170.
XX 01-MAY-1998; 98US-00071739.
XX 04-NOV-1998; 98US-00186200.
XX 19-FEB-2003; 2003US-00368044.
XX
XX (YACO/) YACOBY-ZEEVI O.
XX (PERE/) PERETZ T.
XX (MIRO/) MIRON D.
XX (SHLO/) SHLOMI Y.
XX (PECK/) PECKER I.
XX (AYAL/) AYAL-HERSHKOVITZ M.
XX (FEIN/) FEINSTEIN E.
XX (GELD/) GELDER J M V.
XX (VLOD/) VLODAVSKY I.
XX (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX WPI; 2004-774790/76.
XX
XX New neutralizing monoclonal anti-heparanase antibodies, useful for
XX detecting, treating or preventing cancer, inflammatory or autoimmune
XX disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
XX Claim 67; SEQ ID NO 9; 68pp; English.
XX
XX The invention relates to an isolated antibody or antibody portion capable
XX of specifically binding to or elicited by at least one epitope of a
XX heparanase protein, where the heparanase protein is at least 60%
XX homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
XX where at least one epitope comprises a sequence at least 70% homologous
XX to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
XX a hybridoma cell line comprising a cell line for producing the monoclonal
XX antibody, (b) a method for detecting, treating or preventing a
XX pathological condition or a heparanase-related disorder or condition in a
XX subject, (c) a method for monitoring the state of a heparanase-related
XX disorder or condition in a subject, and (d) a pharmaceutical composition
XX comprising the isolated anti-heparanase antibody or antibody portion and
XX a pharmaceutical carrier. The antibody, methods, and composition are
XX useful for detecting, treating, preventing or monitoring a pathological
XX condition, e.g. angiogenesis, cell proliferation, a cancerous condition
XX (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
XX or prostate cancer), minor cell proliferation, invasion of circulating
XX tumour cells, or a metastatic disease, or a heparanase-related disorder
XX or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
XX (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
XX renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
XX nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
XX carcinoma) in a mammal. This sequence represents a functional peptide
XX epitope of human heparanase.
XX
XX Sequence 15 AA;
Query Match 100.0%; Score 82; DB 8; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;

[illegible]

XX Targeted drug delivery ; inflammatory disorder; wound; scar;
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;
 KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.
 XX
 OS Homo sapiens.
 XX
 XX US2004170631-A1.
 XX
 XX 02-SEP-2004.
 XX
 XX 28-NOV-2003; 2003US-00722502.
 XX
 XX 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 PR 22-AUG-2003; 2003US-00645659.
 XX
 XX (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (VGEL/) VAN GELDER J M.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 PI Friedmann Y;
 XX
 XX WPI; 2004-625084/60.
 XX
 XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.
 XX
 XX Claim 2; SEQ ID NO 1; 58pp; English.
 XX
 XX The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary

CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature
 CC heparanase.
 XX
 XX Sequence 386 AA;
 XX
 XX Query Match 100.0%; Score 82; DB 8; Length 386;
 XX Best Local Similarity 100.0%; Pred. No. 7.9e-05;
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 RFGKKVWLGTSAY 15
 DB 177 RFGKKVWLGTSAY 191
 XX
 RESULT 7
 ADT78174
 ID ADT78174 standard; protein; 386 AA.
 XX
 AC ADT78174;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE 45kDa subunit of mature processed human heparanase dimer.
 XX
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition ; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
 XX
 OS Homo sapiens.
 XX
 XX US2004213789-A1.
 XX
 XX 28-OCT-2004.
 XX
 XX 22-AUG-2003; 2003US-00645659.
 XX
 XX 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX
 XX (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (GELD/) GELDER J M V.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;
 XX
 XX WPI; 2004-774790/76.
 XX
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX
 XX Claim 5; SEQ ID NO 1; 68pp; English.
 PS
 XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of
 CC mature processed human heparanase dimer.

XX
 SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 7.9e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGESSAY 15
 |||||
 DB 177 RPKKKVWLGESSAY 191

RESULT 8

ADY27057
 ID ADY27057 standard; protein; 386 AA.

AC ADY27057;

XX 05-MAY-2005 (first entry)

DE Heparanase inhibitor protein #1.

XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW heparanase modulator; enzyme purification.

XX Homo sapiens.

XX WO2005016227-A2.

XX 24-FEB-2005.

XX 12-AUG-2004; 2004WO-IL000744.

XX 14-AUG-2003; 2003US-0494800P.

XX 12-JAN-2004; 2004US-0535492P.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX Van-Gelder JM, Miron D;

XX WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.

XX Claim 55; SEQ ID NO 33; 211pp; English.

XX

CC The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular
 CC diseases, neurological diseases or viral infections. This sequence
 CC represents a heparanase inhibitor protein used in the scope of the
 CC invention.

XX
 SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;
 Best Local Similarity 100.0%; Pred. No. 7.9e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGESSAY 15
 |||||
 DB 177 RPKKKVWLGESSAY 191

RESULT 9

ADZ18995

ID ADZ18995 standard; protein; 386 AA.

XX ADZ18995;

XX 16-JUN-2005 (first entry)

XX Human heparanase consensus cleavage site #2.

XX Enzyme engineering; heparanase; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.

XX Homo sapiens.

XX WO2005030962-A1.

XX 07-APR-2005.

XX 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

XX 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.

XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.

XX Disclosure; SEQ ID NO 16; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoprotease. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in

CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a human heparanase
 CC consensus cleavage site used in the scope of the invention.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 7.9e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 1 RPQKKVWLGETSSAY 15

|||||

Db 177 RPQKKVWLGETSSAY 191

RESULT 10

AEA42423

ID AEA42423 standard; protein; 386 AA.

XX AC AEA42423;

XX DT 28-JUL-2005 (first entry)

XX DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.

XX KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis.

XX OS Homo sapiens.

XX AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Herschkovitz M, Ben-Artzi H;
 PI Feinstein E;

XX DR WPI; 2005-173343/19.

XX PT Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.

XX PS Claim 2; SEQ ID NO 1; 260pp; English.

XX CC The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides.
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal

CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents the 45 kDa
 CC subunit of the human mature processed heparanase dimer, which is used in
 CC the exemplification of the present invention.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 7.9e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSSAY 15

|||||

Db 177 RPQKKVWLGETSSAY 191

RESULT 11

ADY27061

ID ADY27061 standard; protein; 460 AA.

XX AC ADY27061;

XX DT 05-MAY-2005 (first entry)

XX DE Heparanase inhibitor protein #4.

XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW heparanase modulator; enzyme purification.

XX OS Homo sapiens.

XX WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX PT Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.

XX PS Disclosure; SEQ ID NO 37; 211pp; English.

XX CC The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular

RESULT 14
AEB87562
ID AEB87562 standard; protein; 493 AA.
XX AC AEB87562;
XX DT 06-OCT-2005 (first entry)
XX DE Human heparanase 65delta15 deletion mutant.
XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cystostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; muten.
XX OS Homo sapiens.
OS Synthetic.
XX PN WO2005071070-A2.
XX PD 04-AUG-2005.
XX PF 20-JAN-2005; 2005WO-IL000068.
XX PR 22-JAN-2004; 2004IL-00160025.
XX PR 28-JUL-2004; 2004US-00901943.
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX PI Vlodavsky I, Ilan N, Levy-Adam F;
XX WPI; 2005-564219/57.
XX N-PSDB; AEB87561.
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX Claim 105; SEQ ID NO 4; 167pp; English.
XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX Sequence 493 AA;
Query Match 100.0%; Score 82; DB 9; Length 493;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RPKKKVWLGETSSAY 15
Db 284 RPKKKVWLGETSSAY 298

RESULT 15
ADZ18999
ID ADZ18999 standard; protein; 495 AA.
XX AC ADZ18999;
XX DT 16-JUN-2005 (first entry)
XX DE Hep109 construct protein.
XX KW Enzyme engineering; heparanase; hep109; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX OS Synthetic.
XX PN WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
XX PR 20-JAN-2004; 2004US-0537729P.
XX PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX WPI; 2005-273382/28.
XX N-PSDB; ADZ18998.
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX Example 2; SEQ ID NO 20; 65pp; English.
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hep109 construct protein
CC used in the scope of the invention.
XX Sequence 495 AA;
Query Match 100.0%; Score 82; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RPKKKVWLGETSSAY 15
Db 286 RPKKKVWLGETSSAY 300
RESULT 16
AEB87587
ID AEB87587 standard; protein; 497 AA.
XX AC AEB87587;
XX DT 06-OCT-2005 (first entry)
XX DE Human heparanase 65delta10 deletion mutant.
XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cystostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW

```

KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; mutin.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2005071070-A2.
XX
XX 04-AUG-2005.
XX
XX 20-JAN-2005; 2005WO-IL000068.
XX
XX 22-JAN-2004; 2004IL-00160025.
XX
XX 28-JUL-2004; 2004US-00901943.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Ilan N, Levy-Adam F;
XX
XX WPI; 2005-564219/57.
XX
XX N-PSDB; AEB87586.
XX
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
XX for treating or inhibiting malignant proliferative, inflammatory, kidney
XX disorder or autoimmune disorder.
XX
XX Claim 106; SEQ ID NO 29; 167pp; English.
XX
XX The present sequence is that of a deletion mutant of human heparanase,
XX denoted 65delta10, which is devoid of amino acid residues 270-280 of the
XX native protein. The recombinant protein is deficient of heparanase
XX endoglycosidase catalytic activity. The invention relates to amino acid
XX sequences derived from the N-terminus region of the 50 kDa subunit of
XX heparanase, particularly in the regions between amino acid residues 158-
XX 171, 270-280 and 411-432 of human heparanase. These sequences comprise
XX heparin-binding domains. The invention also provides an antibody directed
XX to these sequences, in particular the 158-171 peptide, and compositions
XX and uses of this antibody as a heparanase inhibitor. A screening method
XX is provided for specific heparanase inhibitors. Claimed pharmaceutical
XX compositions comprising (i) a peptide derived from the N-terminus region
XX of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
XX peptide, or (iii) an antibody which specifically recognizes the peptide
XX are used for the inhibition of heparanase catalytic activity associated
XX with an inflammatory disorder, kidney disease, autoimmune disease,
XX angiogenesis, tumor formation, tumor progression or tumor metastasis, or
XX with a malignant proliferative disorder, especially a solid or non-solid
XX tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
XX Sequence 497 AA;

Query Match          100.0%; Score 82; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPKKKVWLGETSSAY 15
Db 288 RPKKKVWLGETSSAY 302

RESULT 17
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX
XX ADZ19000;
XX
XX 16-JUN-2005 (first entry)
XX
XX HepGS3 construct protein.
XX
XX Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
XX
XX N-PSDB; ADZ19001.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX nucleotide sequence that encodes mammalian heparanase protein and has two
XX consensus cleavage sites located between specific nucleotide encoding
XX residues.
XX
XX Example 2; SEQ ID NO 21; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
XX mammalian heparanase protein, where the nucleic acid comprises two
XX consensus cleavage sites recognized by endoproteinase. The sequences are
XX useful for expressing mammalian heparanase in non-mammalian cells and in
XX inhibitor screening assays for the development of therapeutics or
XX pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
XX and/or inflammation. This sequence represents a hepGS3 construct protein
XX used in the scope of the invention.
XX
XX Sequence 501 AA;

Query Match          100.0%; Score 82; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPKKKVWLGETSSAY 15
Db 292 RPKKKVWLGETSSAY 306

RESULT 18
ADZ19005
ID ADZ19005 standard; protein; 507 AA.
XX
XX ADZ19005;
XX
XX 16-JUN-2005 (first entry)
XX
XX HepGS6 construct protein.
XX
XX Enzyme engineering; heparanase; hepGS6; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
XX

```

PI	Lahm A, Nardella C, Pallaro M, Steinkuhler C;	CC	heparanase activity comprising modulating heparanase activation. The
XX		CC	invention also relates to methods of treating a heparanase- or heparin
DR	WPI; 2005-273382/28.	CC	binding protein-associated disease or disorder in a subject, a
DR	N-PSDB; ADZ19003.	CC	pharmaceutical composition for use in the treatment of a heparanase-
XX		CC	associated disease or disorder comprising a therapeutic amount of an
PT	Synthetic nucleic acid for e.g. inhibitor screening, comprises a	CC	agent capable of modulating heparanase activation and a pharmaceutical
PT	nucleotide sequence that encodes mammalian heparanase protein and has two	CC	carrier or diluent, a method of identifying a protease activator of
PT	consensus cleavage sites located between specific nucleotide encoding	CC	heparanase, a protease substrate mimetic comprising a peptide
PT	residues.	CC	representing a subset or all substrate residues or cleavage sites of
XX		CC	human heparanase or an equivalent non-human heparanase, a method of
PS	Example 2; SEQ ID NO 26; 65pp; English.	CC	producing active heparanase and a method of modulating an adhesion
XX		CC	activity of heparanase. The composition and methods are useful for
CC	The invention relates to a synthetic nucleic acid molecule that encodes	CC	modulating heparanase activation and for treating heparanase-associated
CC	mammalian heparanase protein, where the nucleic acid comprises two	CC	diseases or disorders such as cancer, inflammation, cardiovascular
CC	consensus cleavage sites recognized by endoproteinase. The sequences are	CC	diseases, neurological diseases or viral infections. This sequence
CC	useful for expressing mammalian heparanase in non-mammalian cells and in	CC	represents a human inactive heparanase protein used in the scope of the
CC	inhibitor screening assays for the development of therapeutics or	CC	invention.
CC	pharmaceuticals for inhibiting or treating metastasis, autoimmune disease	XX	
CC	and/or inflammation. This sequence represents a hepgS6 construct protein	SQ	Sequence 508 AA;
CC	used in the scope of the invention.		
SQ	Sequence 507 AA;		
	Query Match 100.0%; Score 82; DB 9; Length 507;		Query Match 100.0%; Score 82; DB 9; Length 508;
	Best Local Similarity 100.0%; Pred. No. 0.00011;		Best Local Similarity 100.0%; Pred. No. 0.00011;
	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 RPKKKVWLGETSSAY 15	Qy	1 RPKKKVWLGETSSAY 15
Db	298 RPKKKVWLGETSSAY 312	Db	299 RPKKKVWLGETSSAY 313
RESULT 19		RESULT 20	
ADY27058		ADZ19006	
ID	ADY27058 standard; protein; 508 AA.	ID	ADZ19006 standard; protein; 526 AA.
XX		XX	
AC	ADY27058;	AC	ADZ19006;
XX		XX	
DT	05-MAY-2005 (first entry)	DT	16-JUN-2005 (first entry)
XX		XX	
DE	Human inactive heparanase protein.	DE	HepHyaluro construct protein.
XX		XX	
KW	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;	KW	Enzyme engineering; heparanase; hepHyaluro; metastasis;
KW	neurological disease; viral infection; infection; cytostatic;	KW	autoimmune disease; inflammation; neoplasm; immune disorder;
KW	antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;	KW	antiinflammatory; cytostatic; immunosuppressive; enzyme.
XX	protease; enzyme; enzyme purification.	XX	
OS	Homo sapiens.	OS	Synthetic.
XX		XX	
PN	WO2005016227-A2.	PN	WO2005030962-A1.
XX		XX	
PD	24-FEB-2005.	PD	07-APR-2005.
XX		XX	
PF	12-AUG-2004; 2004WO-IL000744.	PF	17-SEP-2004; 2004WO-EP010517.
XX		XX	
PR	14-AUG-2003; 2003US-0494800P.	PR	26-SEP-2003; 2003US-0506479P.
XX		XX	
PR	12-JAN-2004; 2004US-0535492P.	PR	20-JAN-2004; 2004US-0537729P.
XX		XX	
PA	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.	PA	(RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX		XX	
PI	Van-Gelder JM, Miron D;	PI	Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX		XX	
DR	WPI; 2005-182203/19.	DR	WPI; 2005-273382/28.
XX		DR	N-PSDB; ADZ19007.
XX		XX	
PT	Regulating heparanase activity, useful for treating heparanase-associated	PT	Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,	PT	nucleotide sequence that encodes mammalian heparanase protein and has two
PT	neurological diseases or viral diseases) comprises modulating heparanase	PT	consensus cleavage sites located between specific nucleotide encoding
PT	activation.	PT	residues.
XX		XX	
PS	Claim 257; SEQ ID NO 34; 21lpp; English.	XX	Example 2; SEQ ID NO 27; 65pp; English.
XX		XX	
CC	The invention relates to a method of regulating heparanase activity in a	CC	The invention relates to a synthetic nucleic acid molecule that encodes
CC	tissue or regulating a biological process depending at least in part on	CC	mammalian heparanase protein, where the nucleic acid comprises two
		CC	consensus cleavage sites recognized by endoproteinase. The sequences are
		CC	useful for expressing mammalian heparanase in non-mammalian cells and in
		CC	inhibitor screening assays for the development of therapeutics or

CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hephyaluro construct
 CC protein used in the scope of the invention.

XX SQ Sequence 526 AA;

Query Match 100.0%; Score 82; DB 9; Length 526;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGESSAY 15
 |||||
 Db 317 RPKKKVWLGESSAY 331

RESULT 21

ABBO7815 ID ABB07815 standard; protein; 527 AA.

XX AC ABB07815;

XX DT 03-JUL-2002 (first entry)

XX DE Chicken signal peptide/human heparanase chimeric protein sequence.

XX KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX OS Synthetic.

OS Gallus gallus.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide /..19

FT Protein /note= "chicken heparanase signal peptide"

FT FT 20..527

XX FT /note= "human heparanase mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2002-338926/37.

XX N-PSDB; ABL40753.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.

XX PS Disclosure; Page 26-28; 39pp; English.

XX CC The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a chicken signal peptide/human heparanase chimeric protein
 CC sequence

SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 5; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGESSAY 15
 |||||
 Db 318 RPKKKVWLGESSAY 332

RESULT 22

ABW02018 ID ABW02018 standard; protein; 527 AA.

XX AC ABW02018;

XX DT 12-FEB-2004 (first entry)

XX DE Chimeric human-chicken heparanase protein.

XX KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
 KW immunosuppressive; vulnary; bactericide; anti-angiogenic; virucide;
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.

XX OS Chimeric - Gallus gallus.

OS Chimeric - Homo sapiens.

XX PN US2003180788-A1.

XX PD 25-SEP-2003.

XX 08-MAY-2003; 2003US-00431438.

XX 20-SEP-2000; 2000US-00666390.

XX 16-AUG-2001; 2001US-00930218.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2003-843931/78.

XX N-PSDB; AAD63532.

XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.

XX Example; Page 26-28; Opp; English.

XX The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
 CC some bacterial and protozoa infections. Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamine. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein

XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 7; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.00011;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGESSAY 15
 Db 318 RPQKKVWLGESSAY 332

RESULT 23
 ADO63825 standard; protein; 527 AA.

XX ADO63825;
 XX 26-AUG-2004 (first entry)
 XX Chimeric heparanase mutant E225A, SEQ ID:10.

XX Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein.

XX Homo sapiens.
 OS Gallus gallus.
 OS Synthetic.
 OS Chimeric.

XX Key Location/Qualifiers
 FT Peptide 1..18
 FT Region 19..527
 FT /note= "Chicken heparanase signal peptide"
 FT /note= "Corresponds to residues 35-543 of human
 FT heparanase mutant E225A (SEQ ID NO:7)"
 FT Misc-difference 209
 FT /note= "Ala replaces wild-type Glu (active site proton
 FT donor). Corresponds to residue 225 of human heparanase
 FT mutant E225A (SEQ ID NO:7)"
 FT Active-site 327
 FT /note= "Active site nucleophile"

XX WO2004048558-A2.
 XX 10-JUN-2004.
 XX 24-NOV-2003; 2003WO-IL000989.
 XX 24-NOV-2002; 2002IL-00153059.
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
 XX WPI; 2004-450373/42.
 XX N-PSDB; ADO63819.

XX New nucleic acid construct comprising heparanase-derived polypeptide,
 useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 thrombasthenia, or Bernard-Soulier syndrome.

XX Claim 10; SEQ ID NO 10; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic
 acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343

are replaced with Ala (ADO63822-ADO63824), and the proteins may
 optionally further comprise an avian heparanase signal peptide (ADO63825-
 ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be use in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase mutant E225A.

XX Sequence 527 AA;
 SQ Query Match 100.0%; Score 82; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGESSAY 15
 Db 318 RPQKKVWLGESSAY 332

RESULT 24
 ADO63825 standard; protein; 527 AA.

XX ADO63825;
 XX 16-JUN-2005 (first entry)
 XX HepGS4 construct protein.

XX Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antinflammatory; cytostatic;
 KW immunosuppressive; enzyme.

XX Synthetic.
 XX WO2005030962-A1.
 XX 07-APR-2005.
 XX 17-SEP-2004; 2004WO-EP010517.
 XX 26-SEP-2003; 2003US-0506479P.
 XX 20-JAN-2004; 2004US-0537729P.
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 XX WPI; 2005-273382/28.
 XX N-PSDB; ADZ19002.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX Example 2; SEQ ID NO 25; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in

CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepgs4 construct protein
 CC used in the scope of the invention.

XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 82; DB 9; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKWLGETSSAY 15
 |||||
 Db 318 RPCKKWLGETSSAY 332

RESULT 25
 AAY34173
 ID AAY34173 standard; protein; 530 AA.
 XX
 AC AAY34173;
 XX
 DT 15-NOV-1999 (first entry)
 XX
 DE Human pre-proheparanase protein sequence.
 XX
 KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9943830-A2.
 XX
 PD 02-SEP-1999.
 XX
 PF 18-FEB-1999; 99WO-US001489.
 XX
 PR 24-FEB-1998; 98US-0075706P.
 PR 26-MAR-1998; 98US-0079401P.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Heinrichson RL, Fairbanks MB, Mildner AM;
 XX
 DR WPI; 1999-540598/45.
 DR N-PSDB; AAZ11236.
 XX
 FT New isolated platelet heparanase polypeptides, used to develop products
 PT for, e.g. wound healing and blocking angiogenesis.
 XX
 PS Claim 12; Fig 7; 57pp; English.
 XX
 CC This sequence is the human pre-proheparanase of the invention. This
 CC sequence was isolated from human platelets. The heparanase can be used
 CC for identifying agents which alter heparanase activity. The heparanase
 CC can be used for wound healing or for blocking angiogenesis or
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic
 CC retinopathy or solid tumours, or for the degradation of heparin and the
 CC neutralisation of heparin's anticoagulant properties during surgery.
 CC Inhibitors of heparanase activity can be used in the treatment of
 CC arthritis, asthma, and other inflammatory diseases, vascular restenosis,
 CC atherosclerosis, tumour growth and progression, fibroproliferative
 CC disorders, and central nervous system (CNS) and neurodegenerative
 CC diseases. The products can also be used for detection and diagnosis. The
 CC purified heparanase, both recombinantly produced human heparanase and
 CC heparanase isolated from human platelet activity, allows for the
 CC convenient selection of compounds having anti-heparanase activity, i.e.
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase
 CC activity. Inhibition of heparanase activity can be measured by blocking

CC heparanase-mediated release of radioactive fragments from in vivo
 CC radiolabelled (HSPG)/heparin
 XX
 SQ Sequence 530 AA;
 Query Match 100.0%; Score 82; DB 2; Length 530;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKWLGETSSAY 15
 |||||
 Db 321 RPCKKWLGETSSAY 335

RESULT 26
 AAY17083
 ID AAY17083 standard; protein; 532 AA.
 XX
 AC AAY17083;
 XX
 DT 21-JUL-1999 (first entry)
 XX
 DE Seq ID No: 15 of WO9921975.
 XX
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
 KW human; HSPG.
 XX
 OS Homo sapiens.
 XX
 PN WO9921975-A1.
 XX
 PD 06-MAY-1999.
 XX
 PF 28-OCT-1998; 98WO-AU000898.
 XX
 PR 28-OCT-1997; 97AU-00000062.
 PR 09-DEC-1997; 97AU-00000812.
 XX
 PA (AUSU) UNIV AUSTRALIAN NAT.
 XX
 PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
 XX
 DR WPI; 1999-312956/26.
 DR N-PSDB; AAX37260.
 XX
 PT Polynucleotides encoding mammalian endoglucuronidases, especially
 PT heparanases, useful to promote wound healing.
 XX
 PS Claim 6; Page 76-79; 112pp; English.
 XX
 CC The invention relates to nucleic acid sequences that encode heparanase
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are
 CC capable of removing the HS side chain from heparan sulfate proteoglycan
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
 CC inhibit heparanase, this is useful for treatment of a physiological or
 CC medical condition associated with elevated heparanase activity, such as
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
 CC rat heparanases can be used to enhance wound healing, especially
 CC associated with tissue development and repair. The conditions mentioned
 CC above can be diagnosed using specific antibodies, and also using primers
 CC and probes specific for the heparanase polynucleotides. Other uses of the
 CC heparanases include sequencing sulfated molecules such as HSPG

XX
 SQ Sequence 532 AA;
 Query Match 100.0%; Score 82; DB 2; Length 532;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKWLGETSSAY 15

```
Db          334  RPCKKWLGETSSAY 348
|||||
RESULT 27
AAV02345
ID AAV02345 standard; protein; 543 AA.
XX
AC AAV02345;
XX
XX AAV02345;
XX
DT 09-JUL-1999 (first entry)
XX
DE A human heparanase protein.
XX
KW Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.
XX
OS Homo sapiens.
XX
XX WO9911798-A1.
XX
XX 11-MAR-1999.
XX
XX 31-AUG-1998; 98WO-US017954.
XX
XX 02-SEP-1997; 97US-00922170.
XX
XX 02-JUL-1998; 98US-00109386.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI; 1999-302255/25.
XX
XX N-PSDB; AAX35648.
XX
XX New human polynucleotide useful for treating angiogenesis, restenosis,
XX and inflammation.
XX
XX Claim 6; Fig 1; 63pp; English.
XX
XX The specification describes a polypeptide having heparanase (hp)
XX activity. The recombinant protein is used as a modulator of heparin-
XX binding growth factors, cellular responses to heparin-binding growth
XX factors and cytokines, cell interaction with plasma lipoproteins,
XX cellular susceptibility to viral, protozoal and bacterial infections or
XX disintegration of neurodegenerative plaques. Heparanase may be useful for
XX conditions such as wound healing, angiogenesis, restenosis,
XX atherosclerosis, inflammation, neurodegenerative diseases, and viral
XX infections. Mammalian heparanase can be used to neutralize plasma
XX heparin, and anti-heparanase antibodies may be applied for
XX immunodetection and diagnosis of micrometastases, autoimmune lesions, and
XX renal failure in biopsy specimens, plasma samples, and body fluids. The
XX present sequence represents human heparanase
XX
XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  RPCKKWLGETSSAY 15
   |||||
Db          334  RPCKKWLGETSSAY 348

RESULT 28
AAV17082
ID AAV17082 standard; protein; 543 AA.
XX
AC AAV17082;
XX
XX 21-JUL-1999 (first entry)
XX
DE Human heparanase enzyme.
XX
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.
XX
XX Homo sapiens.
XX
XX WO9921975-A1.
XX
XX 06-MAY-1999.
XX
XX 28-OCT-1998; 98WO-AU000898.
XX
XX 28-OCT-1997; 97AU-00000062.
XX
XX 09-DEC-1997; 97AU-00000812.
XX
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
XX
XX WPI; 1999-312956/26.
XX
XX N-PSDB; AAX37259.
XX
XX Polynucleotides encoding mammalian endoglucuronidases, especially
XX heparanases, useful to promote wound healing.
XX
XX Claim 6; Page 69-73; 112pp; English.
XX
XX The invention relates to nucleic acid sequences that encode heparanase
XX enzymes having endoglucuronidase activity. Recombinant heparanases are
XX capable of removing the HS side chain from heparan sulfate proteoglycan
XX (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
XX inhibit heparanase, this is useful for treatment of a physiological or
XX medical condition associated with elevated heparanase activity, such as
XX metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
XX arteriosclerosis, atherosclerosis and inflammation. The human, murine and
XX rat heparanases can be used to enhance wound healing, especially
XX associated with tissue development and repair. The conditions mentioned
XX above can be diagnosed using specific antibodies, and also using primers
XX and probes specific for the heparanase polynucleotides. Other uses of the
XX heparanases include sequencing sulfated molecules such as HSPG. The
XX present sequence represents a human heparanase
XX
XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  RPCKKWLGETSSAY 15
   |||||
Db          334  RPCKKWLGETSSAY 348

RESULT 29
AAV57590
ID AAV57590 standard; protein; 543 AA.
XX
AC AAV57590;
XX
XX 02-MAR-2000 (first entry)
XX
DE Human heparanase.
XX
XX Human; heparanase; hpa; genetic modification; expression; anticancer;
```



```

OS Homo sapiens.
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US009255.
XX
PR 01-MAY-1998; 98US-00071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
DR N-PSDB; AAZ33290.
XX
PT Heparanase-specific molecular probes useful for diagnosis and treatment,
PT e.g. of tumors, and for targeted drug delivery.
XX
PS Example; Page 81-82; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and anti-
CC heparanase antibodies are used to detect or quantify the expression of
CC heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
CC sepsis and inflammatory or autoimmune disease). for targeted drug
CC delivery (e.g. of anticancer agents) and as research reagents. The
CC present sequence represents human heparanase, which is used in the
CC exemplification of the present invention
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15
Db |||||
334 RPKKKVWLGETSSAY 348

RESULT 32
AA97635
ID AA97635 standard; protein; 543 AA.
AC AA97635;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL000358.
XX

Query Match 100.0%; Score 82; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15
Db |||||
334 RPKKKVWLGETSSAY 348

RESULT 33
AAB86206
ID AAB86206 standard; protein; 543 AA.
XX
AC AAB86206;
XX
DT 24-AUG-2001 (first entry)
XX
DE Human heparanase inhibitor protein.
XX
KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human.
XX
OS Homo sapiens.
XX
PN DE19955803-A1.
XX
PD 23-MAY-2001.
XX
PF 19-NOV-1999; 99DE-01055803.
XX
PR 19-NOV-1999; 99DE-01055803.
XX
PA (KNOL ) KNOLL AG.
XX
PI Herr D, Hahn A, Laux V;
XX
DR WPI; 2001-368371/39.
DR N-PSDB; AAH20940.
XX
PT Treatment or prevention of cardiac insufficiency and related conditions,
PT e.g. pulmonary congestion and dyspnoea, comprises administration of
PT heparanase inhibitor.
XX
PS Disclosure; Page 11-13; 16pp; German.
XX
CC This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of

```

CC the invention has cardiant, nephrotropic and hepatotropic activity. The
 CC products of the invention can be used in human and veterinary medicine,
 CC for the treatment or prevention of congestive heart failure e.g. primary
 CC cardiomyopathy. Associated conditions treated or prevented with the
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
 CC nocturia can also be treated. This sequence represents the human
 CC heparanase protein described in the method of the invention
 XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 4; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSSAY 15
 |||||
 Db 334 RPQKKVWLGETSSAY 348

RESULT 34
 ABB07813
 ID ABB07813 standard; protein; 543 AA.

XX AC ABB07813;

XX DT 03-JUL-2002 (first entry)

XX DE Human heparanase sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; human.
 XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /note= "signal peptide"
 FT Protein 36..543
 FT /note= "mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;

XX WPI; 2002-338926/37.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.

XX Disclosure; Fig 1a; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a human heparanase protein sequence used in similarity studies

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 5; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSSAY 15
 |||||
 Db 334 RPQKKVWLGETSSAY 348

RESULT 35
 ADD18950
 ID ADD18950 standard; protein; 543 AA.

XX AC ADD18950;

XX DT 15-JAN-2004 (first entry)

XX DE Human disease related protein SeqID439.

XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnery; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transport; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.

XX OS Homo sapiens.

XX WO2003018621-A2.

XX PD 06-MAR-2003.

XX 23-AUG-2002; 2002WO-GB003892.

XX 23-AUG-2001; 2001GB-00020558.

XX 05-OCT-2001; 2001GB-00024037.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX WPI; 2003-290046/28.

XX N-PSDB; ADD18951.

XX New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX Claim 25; SEQ ID NO 439; 424pp; English.

XX This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnery activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011; Mismatches 0; Indels 0; Gaps 0;
Matches 15; Conservative 0;

Qy 1 RPQKKVWLGETSSAY 15
|||||
Db 334 RPQKKVWLGETSSAY 348

RESULT 36

ADG88800
ID ADG88800 standard; protein; 543 AA.

AC ADG88800;

XX 11-MAR-2004 (first entry)

XX Human hpa protein.

KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.

XX Homo sapiens.

XX US2003161823-A1.

XX 28-AUG-2003.

XX 14-JAN-2003; 2003US-00341582.

XX 31-AUG-1998; 98WO-US017954.

XX 01-MAR-1999; 99US-00258892.

XX 06-FEB-2001; 2001US-00776874.

XX 05-SEP-2001; 2001WO-IL000830.

XX 19-NOV-2001; 2001US-00988113.

XX (ILAN/) ILAN N.

XX (VLOD/) VLODAVSKY I.

XX (YACO/) YACOBY-ZEEVI O.

XX (PECK/) PECKER I.

XX (FEIN/) FEINSTEIN E.

XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;

XX WPI; 2003-897910/82.

XX N-PSDB; ADG88799, ADG88801, ADG88832.

XX Composition for treating a wound comprising recombinant heparanase is
XX useful to induce or accelerate wound healing and induce or accelerate
XX angiogenesis.

XX Claim 2; SEQ ID NO 10; 143pp; English.

XX The present invention relates to methods and compositions for inducing
XX and/or accelerating wound healing via the catalytic activity of
XX heparanase. The invention is used to induce or accelerate a healing
XX process, particularly of an ulcer, burn, laceration, surgical incision,
XX necrosis, pressure wound, diabetic ulcer and to induce or accelerate
XX angiogenesis. The present sequence is human hpa protein.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 7; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.00011; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

Qy 1 RPQKKVWLGETSSAY 15
|||||

Db 334 RPQKKVWLGETSSAY 348

RESULT 37

ADL16379

ID ADL16379 standard; protein; 543 AA.

XX ADL16379;

XX 06-MAY-2004 (first entry)

XX Human heparanase partial protein.

XX Human; heparanase; heparanase-dependent cancer; cancer;
XX autoimmune reaction; inflammation; chromosome 4; enzyme.

XX Homo sapiens.

XX US2003236215-A1.

XX 25-DEC-2003.

XX 09-JUN-2003; 2003US-00456573.

XX 31-AUG-1998; 98WO-US017954.

XX 01-MAR-1999; 99US-00258892.

XX 08-NOV-1999; 99US-00435739.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Pecker I, Vlodavsky I, Feinstein E;

XX WPI; 2004-070610/07.

XX New antisense oligonucleotide hybridizable with a polynucleotide encoding
XX a polypeptide with heparanase activity, useful for treating diseases such
XX as cancer and autoimmune disorders.

XX Claim 3; SEQ ID NO 10; 108pp; English.

XX The invention relates to an antisense oligonucleotide (ASO) comprising a
XX polynucleotide or a polynucleotide analogue of at least 10 bases being
XX hybridisable in vivo, under physiological conditions, with a portion of
XX a polynucleotide strand encoding a polypeptide having heparanase
XX catalytic activity. Also included are a method of in vivo downregulating
XX heparanase activity (comprising administering the ASO in vivo), a method
XX of treating a subject suffering from a pathological condition
XX (characterised by heparanase activity, comprising administering ASO to
XX the subject), a pharmaceutical composition comprising the ASO and a
XX carrier, an antisense nucleic acid construct (comprising a promoter
XX sequence and a polynucleotide sequence directing the synthesis of an
XX antisense RNA sequence of at least 10 bases being hybridisable in vivo,
XX under physiological conditions, with a polynucleotide strand encoding a
XX polypeptide having heparanase catalytic activity), a method of in vivo
XX downregulating heparanase activity (comprising administering in vivo the
XX antisense nucleic acid construct), a pharmaceutical composition
XX comprising the antisense nucleic acid construct and a carrier, and an
XX antisense oligonucleotide comprising a polynucleotide or a polynucleotide
XX analogue of at least 10 bases being hybridisable in vivo, under
XX physiological conditions, with a portion of a polynucleotide strand being
XX characterised by forming at least a portion of an untranslated region
XX (UTR) for a polynucleotide strand encoding a polypeptide having
XX heparanase catalytic activity. The methods and compositions of the
XX present invention are useful for the prevention and/or treatment of
XX diseases or conditions associated with aberrant heparanase activity, such
XX as heparanase-dependent cancer, cancer, autoimmune reaction and
XX inflammation. The gene for human heparanase is located on chromosome 4.

XX The present sequence is a human heparanase protein.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.00011; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

Qy 1 RPQKKVWLGETSSAY 15
|||||

```
Db      334  RPKKVLGETSSAY 348

RESULT 38
ADK52086
ID  ADK52086 standard; protein; 543 AA.
XX
AC  ADK52086;
XX
XX  20-MAY-2004 (first entry)
XX
XX  Human atopic dermatitis/psoriasis-associated protein #1.
XX
XX  Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;
KW  antipsoriatic; rash.
XX
XX  Homo sapiens.
XX
XX  W02004016785-A1.
XX
XX  26-FEB-2004.
XX
XX  06-AUG-2003; 2003WO-JP009999.
XX
XX  06-AUG-2002; 2002JP-00229319.
XX
XX  14-MAY-2003; 2003JP-00136544.
XX
XX  (GENO-) GENOX RES INC.
XX
XX  (UJGU-) UNIV JUNTENDO.
XX
XX  Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
PI  Mitsuishi K;
XX
XX  WPI; 2004-214514/20.
XX
XX  N-PSDB; ADK51968.
XX
XX  Detecting atopic dermatitis or psoriasis comprises assaying levels of
PT  expression of an indicator gene at a rash site and non-rash site of a
PT  person with atopic dermatitis or psoriasis.
XX
XX  Example 2; SEQ ID NO 119; 484pp; Japanese.
XX
XX  The invention relates to detecting atopic dermatitis or psoriasis
CC  comprising assaying the levels of expression of an indicator gene at a
CC  rash site and non-rash site of a person with atopic dermatitis or
CC  psoriasis, comparing these levels with those of a healthy person, and
CC  determining that if the levels of indicators are higher or lower, then
CC  this indicates the disease. Also included are a reagent for detecting
CC  atopic dermatitis or psoriasis, a kit for screening for treatments, a
CC  transgenic non human vertebrate animal models for the diseases, an agent
CC  for inducing the diseases in mice and a DNA chip for assaying for the
CC  indicator genes. The method is used for treatment, detection and animal
CC  models for research of atopic dermatitis and psoriasis. The present
CC  sequence is a protein encoded by an indicator gene of the invention.
XX
XX  Sequence 543 AA;
SQ

Query Match      100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  RPKKVLGETSSAY 15
      |||||
Db      334  RPKKVLGETSSAY 348

RESULT 39
ADM48716
ID  ADM48716 standard; protein; 543 AA.
XX
AC  ADM48716;
XX
XX  03-JUN-2004 (first entry)
XX
XX  Human hpa protein #2.
```

```
XX
DE      Human hpa protein #1.
XX
XX  Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW  neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW  human.
XX
XX  Homo sapiens.
XX
XX  US2003217375-A1.
XX
XX  20-NOV-2003.
XX
XX  24-FEB-2003; 2003US-00371218.
XX
XX  31-AUG-1998; 98WO-US017954.
XX
XX  01-MAR-1999; 99US-00258892.
XX
XX  06-FEB-2001; 2001US-00776874.
XX
XX  19-NOV-2001; 2001US-00988113.
XX
XX  (ZCHA/) ZCHARIA E.
XX
XX  (VLOD/) VLODAVSKY I.
XX
XX  (METZ/) METZGER S.
XX
XX  (PECK/) PECKER I.
XX
XX  (ILAN/) ILAN N.
XX
XX  (CHAJ/) CHAJEK-SHAUL T.
XX
XX  (GOLD/) GOLDSHMIDT O.
XX
XX  Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
PI  Chajek-Shaul T, Goldshmidt O;
XX
XX  WPI; 2004-021918/02.
XX
XX  N-PSDB; ADM48715, ADM48717.
XX
XX  New transgenic non-human animal expressing heparinase, useful as models
PT  for human disease, such as cancers, viral infection, neurodegenerative
PT  diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
XX  Example 1; SEQ ID NO 10; 106pp; English.
XX
XX  The present invention relates to a transgenic non-human animal whose
CC  genome comprises an exogenous polynucleotide sequence, including a
CC  promoter active in tissues of the non-human, a region encoding a human
CC  heparanase, where the promoter and the region encoding human heparanase
CC  are operably linked in the exogenous polynucleotide such that human
CC  heparanase is expressed in at least a portion of the cells of the non-
CC  human animal. The methods and compositions of the present invention are
CC  useful for the production of transgenic animals expressing heparanase, to
CC  be used as models for human diseases such as cancers, viral infection,
CC  restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC  disorders. The present sequence is human hpa protein used in the
CC  exemplification of the invention.
XX
XX  Sequence 543 AA;
SQ

Query Match      100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  RPKKVLGETSSAY 15
      |||||
Db      334  RPKKVLGETSSAY 348

RESULT 40
ADM48759
ID  ADM48759 standard; protein; 543 AA.
XX
AC  ADM48759;
XX
XX  03-JUN-2004 (first entry)
XX
XX  Human hpa protein #2.
```

XX Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW human.
XX
XX Homo sapiens.
XX OS
PN US2003217375-A1.
XX
XX 20-NOV-2003.
PD
XX
XX 24-FEB-2003; 2003US-00371218.
XX
XX 31-AUG-1998; 98WO-US017954.
PR
XX 01-MAR-1999; 99US-00258892.
PR
XX 06-FEB-2001; 2001US-00776874.
PR
XX 19-NOV-2001; 2001US-00988113.
XX
XX (ZCHA/) ZCHARIA E.
PA (VLOD/) VLODAVSKY I.
PA (METZ/) METZGER S.
PA (PECK/) PECKER I.
PA (ILAN/) ILAN N.
PA (CHAJ/) CHAJEK-SHAUL T.
PA (GOLD/) GOLDSHMIDT O.
XX
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX
XX WPI: 2004-021918/02.
DR
XX N-PSDB; ADM48748.
DR
XX
XX New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
XX Example 10; Fig 16; 106pp; English.
PS
XX
XX The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPKKKVWLGETSSAY 15
DB 334 RPKKKVWLGETSSAY 348
RESULT 41
ADN05074
ID ADN05074 standard; protein; 543 AA.
XX
XX ADN05074;
AC
XX 01-JUL-2004 (first entry)
DT
XX Antipsoriatic protein sequence #716.
DE
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
KW

XX Homo sapiens.
XX OS
PN WO2004028479-A2.
XX
XX 08-APR-2004.
PD
XX
XX 25-SEP-2003; 2003WO-US030907.
PR
XX 25-SEP-2002; 2002US-0414006P.
PR
XX (GETH) GENENTECH INC.
PA
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
XX WPI: 2004-305105/28.
DR
XX N-PSDB; ADN05073.
DR
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
XX Claim 9; SEQ ID NO 1468; 3069pp; English.
PS
XX
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPKKKVWLGETSSAY 15
DB 334 RPKKKVWLGETSSAY 348
RESULT 42
ADN04902
ID ADN04902 standard; protein; 543 AA.
XX
XX ADN04902;
AC
XX 01-JUL-2004 (first entry)
DT
XX Antipsoriatic protein sequence #631.
DE
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
KW
XX Homo sapiens.
XX OS
PN WO2004028479-A2.
XX
XX 08-APR-2004.
PD
XX
XX 25-SEP-2003; 2003WO-US030907.
PR
XX 25-SEP-2002; 2002US-0414006P.
PR
XX (GETH) GENENTECH INC.
PA
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
XX WPI: 2004-305105/28.
DR
XX N-PSDB; ADN04901.
DR
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT antipsoriatic; gene therapy; psoriasis; diagnosis.
KW

PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 XX
 PS Claim 9; SEQ ID NO 1296; 3069pp; English.
 XX
 CC The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polypeptides of the invention.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RPQKKVWLGETSSAY 15
 DB 334 RPQKKVWLGETSSAY 348
 |||||
 RESULT 43
 ADO63831
 ID ADO63831 standard; protein; 543 AA.
 XX
 AC ADO63831;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human heparanase mutant E378A.
 XX
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein; enzyme.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Active-site 225 /note= "Active site proton donor"
 FT Active-site 343 /note= "Active site nucleophile"
 FT Misc-difference 378 /note= "Ala replaces wild-type Glu"
 FT
 FT
 FT
 FN WO2004048558-A2.
 XX
 XX 10-JUN-2004.
 XX
 XX 24-NOV-2003; 2003WO-IL000989.
 XX
 XX 24-NOV-2002; 2002IL-00153059.
 XX
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
 XX WPI; 2004-450373/42.
 XX
 XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 XX
 XX Example 4; Page; 128pp; English.
 PS
 XX The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase

CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a human
 CC heparanase mutant E378A created in an example of the invention which
 CC retains its heparanase catalytic activity. The present sequence is not
 CC shown in the invention, but is derived from the protein sequence of
 CC GenBank accession number AF144325 and the information provided on page
 CC 70.
 XX Sequence 543 AA;
 SQ
 Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RPQKKVWLGETSSAY 15
 DB 334 RPQKKVWLGETSSAY 348
 |||||
 RESULT 44
 ADO63832
 ID ADO63832 standard; protein; 543 AA.
 XX
 AC ADO63832;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human heparanase mutant E396A.
 XX
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein; enzyme.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Active-site 225 /note= "Active site proton donor"
 FT Active-site 343 /note= "Active site nucleophile"
 FT Misc-difference 396 /note= "Ala replaces wild-type Glu"
 FT
 FT
 FT
 FN WO2004048558-A2.
 XX
 XX 10-JUN-2004.
 XX

ADQ80372
ID ADQ80372 standard; protein; 543 AA.
XX
AC ADQ80372;
XX
DT 21-OCT-2004 (first entry)
XX
DE Heparanase protein.
XX
KW cytostatic; epidermal growth factor receptor modulator; identification;
KW therapeutic response; cancer; EGFR; biomarker.
XX
OS Homo sapiens.
XX
PN WO2004063709-A2.
XX
PD 29-JUL-2004.
XX
XX 08-JAN-2004; 2004WO-US000368.
PF
XX 08-JAN-2003; 2003US-0438735P.
PR
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
PI Amler LC, Januario T;
XX
XX WPI; 2004-544114/52.
DR N-PSDB; ADQ80253.
XX
XX Identifying a mammal that will respond therapeutically to a method of
PT treating cancer comprises comparing the level of a biomarker in a mammal
PT before and after exposure to an epidermal growth factor receptor (EGFR)
PT modulator.
XX
PS Disclosure; SEQ ID NO 144; 520pp; English.
XX
CC The invention relates to a method of identifying a mammal that will
CC respond therapeutically to a method of treating cancer by administering
CC an epidermal growth factor receptor (EGFR) modulator by comparing the
CC level of a biomarker in a mammal before and after exposure to an EGFR
CC modulator. The method comprises: (a) measuring, in the mammal, the level
CC of at least one biomarker identified in the specification; (b) exposing
CC the mammal to the EGFR modulator; and (c) measuring in the mammal the
CC level of the biomarker, where a difference in the level in step (c)
CC compared to step (a) indicates that the mammal will respond
CC therapeutically to the method of treating cancer. The method and
CC biomarkers are useful for identifying a mammal that will respond
CC therapeutically to a method of treating cancer by administering an
CC epidermal growth factor receptor (EGFR) modulator. This sequence
CC corresponds to one of the biomarkers whose levels of expression is
CC measured in the method of the invention.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKGKWLGETSSAY 15
|||||
Db 334 RPKGKWLGETSSAY 348

RESULT 47
ADR88210
ID ADR88210 standard; protein; 543 AA.
XX
AC ADR88210;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human preproheparanase.
XX

Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
autoimmune disorder; cancer; angiogenesis; metastatic disease;
atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
human; heparanase; enzyme.
Homo sapiens.
Key Location/Qualifiers
Peptide 1..35
Protein /label= Signal_peptide
36..543
Region /label= Mature_heparanase
36..109
Domain /note= "8 KDa subunit of mature heparanase dimer"
89..107
FT /note = Functional peptide epitope
158..543
FT /note= "45 KDa subunit of mature heparanase dimer"
219..233
FT /note = Functional peptide epitope
225
FT /note= "Active site residue"
258..266
FT /note= "Putative heparin binding domain"
294..307
FT /note = Functional peptide epitope
334..348
FT /note = Functional peptide epitope
343
FT /note= "Active site residue"
414..420
FT /note= "Putative heparin binding domain"
437..446
FT /note = Functional peptide epitope
US2004170631-A1.
02-SEP-2004.
28-NOV-2003; 2003US-00722502.
02-SEP-1997; 97US-00922170.
01-MAY-1998; 98US-00071739.
04-NOV-1998; 98US-00186200.
19-FEB-2003; 2003US-00368044.
22-AUG-2003; 2003US-00645659.
(YACO/) YACOBY-ZEEVI O.
(PERE/) PERETZ T.
(PA (MIRO/) MIRON D.
(PA (SHLO/) SHLOMI Y.
(PA (PECK/) PECKER I.
(PA (AYAL/) AYAL-HERSHKOVITZ M.
(PA (FEIN/) FEINSTEIN E.
(PA (VGEL/) VAN GELDER J M.
(PA (VLOD/) VLODAVSKY I.
(PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
DR WPI; 2004-625084/60.
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX
PS Claim 2; SEQ ID NO 4; 58pp; English.
XX

CC The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an epitope of its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is human preproheparanase.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15
 |||||
 Db 334 RPQKKVWLGETSSAY 348

RESULT 48

ADP25079
 ID ADP25079 standard; protein; 543 AA.

XX AC ADP25079;

XX 18-NOV-2004 (first entry)

XX PRO polypeptide SEQ ID NO:2257.

XX PRO; antiinflammatory; antiarthritic; antirheumatic; immunosuppressive;
 KW osteopathic; antidiabetic; dermatological; antipsoriatic; antiallergic;
 KW antiasthmatic; hepatotropic; respiratory; gene therapy; immune system.

OS Unidentified.

XX WO2004041170-A2.

XX 21-MAY-2004.

XX 30-OCT-2003; 2003WO-US034312.

XX 01-NOV-2002; 2002US-0423394P.

XX (GETH) GENENTECH INC.

XX Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WI;
 PI Wu TD;

DR WPI: 2004-419628/39.
 DR N-PSDB; ADP25078.

XX New PRO polypeptides and polynucleotides, useful for treating e.g.
 PT erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated
 PT renal disease, or demyelinating diseases of the central or peripheral
 PT nervous system.

XX Claim 7; SEQ ID NO 2257; 2940pp; English.

XX The invention relates to a novel isolated nucleic acid and the PRO
 CC polypeptide encoded by it. A protein of the invention has
 CC antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,
 CC osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,
 CC antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide
 CC of the invention may have a use in gene therapy. The PRO polypeptide, its
 CC agonist, antagonist, or antibody that specifically binds to the
 CC polypeptide is useful for treating an immune related disorder such as
 CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,
 CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an
 CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic
 CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune
 CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal
 CC disease, a demyelinating disease of the central or peripheral nervous
 CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,
 CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary
 CC disease, infectious or autoimmune chronic active hepatitis, primary
 CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,
 CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's
 CC disease, an autoimmune or immune-mediated skin disease, a bullous skin
 CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic
 CC disease, asthma, allergic rhinitis, atopic dermatitis, food
 CC hypersensitivity, urticaria, an immunologic disease of the lung,
 CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity
 CC pneumonitis, a transplantation associated disease, graft rejection or
 CC graft-versus-host disease. The present sequence represents a PRO protein
 CC of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15
 |||||
 Db 334 RPQKKVWLGETSSAY 348

RESULT 49

ADT78177
 ID ADT78177 standard; protein; 543 AA.

XX AC ADT78177;

XX 13-JAN-2005 (first entry)

XX Human heparanase protein.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Binding-site 157..162

FT Binding-site /note= "Putative heparin binding site"

FT Binding-site 271..277

FT Binding-site /note= "Putative heparin binding site"

FT Binding-site 426..433
XX /note= "Putative heparin binding site"
XX US2004213789-A1.
XX
XX
XX
XX 28-OCT-2004.
XX
XX 22-AUG-2003; 2003US-00645659.
XX
XX 02-SEP-1997; 97US-00922170.
XX 01-MAY-1998; 98US-00071739.
XX 04-NOV-1998; 98US-00186200.
XX 19-FEB-2003; 2003US-00368044.
XX
XX (YACO/) YACOBY-ZEEVI O.
XX (PERE/) PERETZ T.
XX (MIRO/) MIRON D.
XX (SHLO/) SHLOMI Y.
XX (PECK/) PECKER I.
XX (AYAL/) AYAL-HERSHKOVITZ M.
XX (FEIN/) FEINSTEIN E.
XX (GELD/) GELDER J M V.
XX (VLOB/) VLODAVSKY I.
XX (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI; 2004-774790/76.
XX
XX New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
XX Claim 5; SEQ ID NO 4; 68pp; English.
XX
XX The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents human heparanase.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPKKKVWLGETSSAY 15
DB 334 RPKKKVWLGETSSAY 348
RESULT 50

ADY27036
ID ADY27036 standard; protein; 543 AA.
XX
XX AC ADY27036;
XX
XX DT 05-MAY-2005 (first entry)
XX
XX DE Human heparanase protein.
XX
XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW protease; enzyme; enzyme purification.
XX
XX OS Homo sapiens.
XX
XX PN WO2005016227-A2.
XX
XX PD 24-FEB-2005.
XX
XX PF 12-AUG-2004; 2004WO-IL000744.
XX
XX PR 14-AUG-2003; 2003US-0494800P.
XX
XX PR 12-JAN-2004; 2004US-0535492P.
XX
XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX
XX PI Van-Gelder JM, Miron D;
XX
XX DR WPI; 2005-182203/19.
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.
XX
XX Disclosure; SEQ ID NO 8; 21pp; English.
XX
XX The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a human heparanase protein used in the scope of the invention.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RPKKKVWLGETSSAY 15
DB 334 RPKKKVWLGETSSAY 348
Search completed: June 5, 2006, 12:41:01
Job time : 136.384 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 16.6438 Seconds

(without alignments)
86.714 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPCKVWLGETSSAY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: PIR_80.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	57.3	426	2 S44953	lmbf protein - Str
2	46	56.1	1180	1 NCECX5	exodeoxyribonuclea
3	46	56.1	1180	2 G85933	DNA helicase RecB
4	46	56.1	1180	2 E91088	DNA helicase RecB
5	45.5	55.5	1596	2 AG2501	hypothetical prote
6	45	54.9	177	2 D84828	AP2 domain transcr
7	45	54.9	562	2 AD0783	PTS system, fructo
8	44	53.7	336	2 E84594	AP2 domain transcr
9	44	53.7	563	2 E85855	PTS system, fructo
10	44	53.7	563	2 C91011	fructose-specific
11	44	53.7	563	2 A34962	phosphotransferase
12	43	52.4	249	2 A83664	tRNA/rRNA methyltr
13	43	52.4	1022	2 T51257	calmodulin-binding
14	43	52.4	1022	2 T50928	calmodulin-binding
15	43	52.4	1181	2 AB0865	exonuclease V chai
16	42	51.2	283	2 C84828	AP2 domain transcr
17	42	51.2	366	2 C83034	probable oxidoredu
18	42	51.2	775	2 E70320	polyribonucleotide
19	42	51.2	1366	1 CGH2S	collagen alpha 2(I
20	41	50.0	225	2 T02433	DNA binding protei
21	41	50.0	306	2 AH2866	S-methyltransferas
22	41	50.0	306	2 E97643	mesh protein limpor
23	41	50.0	420	2 S36444	hygromycin phospho
24	41	50.0	480	2 JC7506	heparanase protein
25	41	50.0	615	2 AH2248	proteinase limpor
26	41	50.0	1220	2 AD0125	exodeoxyribonuclea
27	40	48.8	122	2 AE3301	2-deoxy-D-gluconat
28	40	48.8	131	2 S64937	probable membrane
29	40	48.8	222	2 T52019	ethylene responsiv

30	40	48.8	239	2 G97553	sugar fermentation
31	40	48.8	239	2 A12773	sugar fermentation
32	40	48.8	243	2 B82979	hypothetical prote
33	40	48.8	276	2 E72623	probable autoantig
34	40	48.8	291	2 T02434	DNA binding protei
35	40	48.8	326	2 T09544	phosphoprotein pho
36	40	48.8	386	2 JC5191	contractile tail s
37	40	48.8	958	2 S73012	polyketide synthas
38	40	48.8	1014	2 F95501	hypothetical prote
39	40	48.8	1126	1 WPFM12	125K protein - alf
40	40	48.8	1871	2 AB7204	polyketide synthas
41	39	47.6	115	2 AB2718	conserved hypothet
42	39	47.6	182	2 F97499	hypothetical prote
43	39	47.6	184	2 A86461	hypothetical prote
44	39	47.6	196	2 H64968	acetyl CoA acetyl
45	39	47.6	221	2 T17997	hypothetical prote
46	39	47.6	308	2 F64901	ABC-type transport
47	39	47.6	308	2 B85728	hypothetical prote
48	39	47.6	308	2 G90889	hypothetical prote
49	39	47.6	366	1 XNB7UG	UDPglucose-hexose-
50	39	47.6	379	2 B70579	probable cell divi
51	39	47.6	382	1 SUBSN	subtilisin (EC 3.4
52	39	47.6	410	2 F70004	probable mannose-6
53	39	47.6	425	2 F70038	arabinogalactan en
54	39	47.6	751	2 G01234	N-ethylmaleimide-s
55	39	47.6	752	2 S04235	vesicular fusion p
56	39	47.6	828	2 T06133	hypothetical prote
57	39	47.6	1197	2 D82696	hypothetical prote
58	39	47.6	1373	1 A43291	collagen alpha 2(I
59	39	47.6	1760	2 E86201	protein Fl2K11.4 [
60	39	47.6	4151	2 G70944	probable polyketid
61	38	46.3	136	2 T36624	hypothetical prote
62	38	46.3	143	2 S07234	collagen alpha 1(I
63	38	46.3	183	2 T14653	hypothetical prote
64	38	46.3	195	2 T14965	hypothetical prote
65	38	46.3	204	2 S49031	cadmium-induced pr
66	38	46.3	207	2 D82423	hypothetical prote
67	38	46.3	248	2 T07728	transcription fact
68	38	46.3	251	2 T01986	Tsai protein - com
69	38	46.3	281	2 T52189	ethylene responsiv
70	38	46.3	297	2 G70185	nevalonate kinase
71	38	46.3	323	2 F90479	(Phospho) nevalona
72	38	46.3	323	2 T51833	transcription fact
73	38	46.3	358	2 D96579	hypothetical prote
74	38	46.3	362	2 F75379	S-adenosylmethioni
75	38	46.3	378	2 S54056	probable membrane
76	38	46.3	437	2 S60957	transcription modu
77	38	46.3	444	2 C83336	probable cytochrom
78	38	46.3	454	1 NMIV	exo-alpha-sialidas
79	38	46.3	462	2 S10235	alpha-L-fucosidase
80	38	46.3	470	1 NMIVAK	exo-alpha-sialidas
81	38	46.3	470	1 NMIVXL	exo-alpha-sialidas
82	38	46.3	470	2 S04801	exo-alpha-sialidas
83	38	46.3	473	2 S10629	collagen - chicken
84	38	46.3	493	2 G83564	probable ATPase PA
85	38	46.3	520	2 S74497	hypothetical prote
86	38	46.3	603	2 E71444	probable EREBP-4 -
87	38	46.3	613	2 JC7827	X-Pro aminopeptida
88	38	46.3	662	2 T04461	potassium channel
89	38	46.3	810	2 T10756	Nel-homolog protei
90	38	46.3	958	2 S41013	hypothetical prote
91	38	46.3	1010	2 T09499	ATP-dependent clp
92	38	46.3	1032	2 T23164	hypothetical prote
93	38	46.3	1035	2 T23165	hypothetical prote
94	38	46.3	1040	2 D88568	protein ZK57.3 [1
95	38	46.3	3140	1 GNV5RA	genome polyprotein
96	38	46.3	3140	2 S47508	genome polyprotein
97	37.5	45.7	279	2 AC3114	hypothetical prote
98	37.5	45.7	304	2 B98173	hypothetical prote
99	37.5	45.7	1004	2 T38074	hypothetical prote
100	37.5	45.7	1401	2 T17452	Werner syndrome pr
101	37.5	45.7	2581	2 AF2545	hypothetical prote
102	37	45.1	94	2 T15982	hypothetical prote

103	37	45.1	161	2	T07689	transcription fact	176	36	43.9	300	2	T52020	ethylene responsiv
104	37	45.1	196	2	H44701	ribosomal protein	177	36	43.9	306	2	B28170	reaction center pr
105	37	45.1	206	2	S75258	probable 3-demethy	178	36	43.9	310	2	F81878	probable prolyl am
106	37	45.1	258	2	A13450	ferredoxin-NADP re	179	36	43.9	310	2	B81141	proline iminopecti
107	37	45.1	270	2	C86670	hypothetical prote	180	36	43.9	317	2	T03304	probable phosphopr
108	37	45.1	303	2	T04541	hypothetical prote	181	36	43.9	318	2	D82742	dihydroxydipicolin
109	37	45.1	344	2	T44032	integral membrane	182	36	43.9	323	2	A32549	phosphoprotein pho
110	37	45.1	364	2	D75539	branched-chain ami	183	36	43.9	325	2	T09547	phosphoprotein pho
111	37	45.1	396	2	AF1903	hypothetical prote	184	36	43.9	325	2	AH2555	hypothetical prote
112	37	45.1	408	2	G87646	molybdopterin bios	185	36	43.9	326	2	T09995	phosphoprotein pho
113	37	45.1	422	2	D84403	dihydrocorotase lim	186	36	43.9	334	2	D84978	lipoprotein nlpD p
114	37	45.1	428	2	T24769	hypothetical prote	187	36	43.9	337	2	E86543	probable RAV-like
115	37	45.1	432	2	AC0161	serine-type D-Ala-	188	36	43.9	380	2	G89604	spore coat protein
116	37	45.1	441	2	H82642	hypothetical prote	189	36	43.9	402	2	F81387	UDP-N-acetylmuramo
117	37	45.1	443	1	BVB91	mevalonate kinase	190	36	43.9	407	1	KHRTD	cathepsin D (EC 3.
118	37	45.1	451	2	G01227	recepin - human	191	36	43.9	407	2	T36649	probable transposa
119	37	45.1	455	2	AD2049	aminopeptidase P [192	36	43.9	408	2	H95416	probable ROK famil
120	37	45.1	468	1	NMIVAA	exo-alpha-stalidas	193	36	43.9	410	1	KHMSD	cathepsin D (EC 3.
121	37	45.1	512	2	A55206	sucrose/fructan hy	194	36	43.9	412	1	KHHUD	cathepsin D (EC 3.
122	37	45.1	521	2	T45608	hypothetical prote	195	36	43.9	416	2	D84829	hypothetical prote
123	37	45.1	521	2	A70954	hypothetical prote	196	36	43.9	442	2	JC5218	type I site-specif
124	37	45.1	579	2	G69252	aldehyde ferredoxi	197	36	43.9	447	2	S20711	neuraminidase - in
125	37	45.1	592	2	E89772	hypothetical prote	198	36	43.9	453	1	NMIV3	exo-alpha-sialidas
126	37	45.1	649	2	T33741	DNA-binding protei	199	36	43.9	466	2	S37113	nitric-oxide reduc
127	37	45.1	663	1	A69798	beta-galactosidase	200	36	43.9	470	1	NMIVEK	exo-alpha-sialidas
128	37	45.1	703	1	WMBE7	Uil17 protein - hum	201	36	43.9	498	2	H81782	adhesin MafB2 NMA2
129	37	45.1	722	2	B75608	GWC oxidoreductase	202	36	43.9	501	2	C82414	aminopeptidase VCA
130	37	45.1	735	2	JC5869	beta-glucosidase (203	36	43.9	506	2	T12819	hypothetical prote
131	37	45.1	846	2	T19049	hypothetical prote	204	36	43.9	533	2	D86756	prophage p12 prote
132	37	45.1	825	2	H64083	biotin sulfoxide r	205	36	43.9	587	1	TVFVPR	protein-tyrosine k
133	37	45.1	841	2	AG2250	nitrogen assimilat	206	36	43.9	667	2	E86728	NADH dehydrogenase
134	37	45.1	903	2	C83044	Mg(2+) transport A	207	36	43.9	686	2	T12967	hypothetical prote
135	37	45.1	1556	1	A60988	saliva-interacting	208	36	43.9	711	2	C83922	ATP-dependent prot
136	37	45.1	1561	1	S06839	surface antigen sp	209	36	43.9	812	2	H86265	protein F3P19.18 l
137	37	45.1	1585	2	S04729	hypothetical prote	210	36	43.9	824	2	T36818	probable secreted
138	37	45.1	1630	2	T40217	microtubule bindin	211	36	43.9	825	2	A59296	alpha-L-arabinofur
139	37	45.1	1690	2	T13030	calcium channel pr	212	36	43.9	854	2	C83905	leucyl-tRNA synthet
140	37	45.1	2178	2	S29237	phosphoribosyl-AMP	213	36	43.9	872	2	AD2216	leucyl-tRNA synthet
141	36.5	44.5	85	2	A90207	Ig heavy chain V-I	214	36	43.9	915	2	S54485	CESI protein - yea
142	36.5	44.5	86	2	F34964	probable acyl-coad	215	36	43.9	928	2	AF3516	Mg(2+) transport A
143	36.5	44.5	319	2	T70605	hypothetical prote	216	36	43.9	950	2	F86286	hypothetical prote
144	36.5	44.5	331	2	A35809	probable membrane	217	36	43.9	1039	2	T30856	protein F2 - Strept
145	36.5	44.5	380	2	D90749	probable membrane	218	36	43.9	1249	2	AC1065	helicase related p
146	36.5	44.5	380	2	H85599	probable membrane	219	36	43.9	1301	2	S18118	alpha-amylase - Al
147	36.5	44.5	380	2	F64826	probable membrane	220	36	43.9	1404	2	F86470	probable retroelem
148	36.5	44.5	907	2	F84825	hypothetical prote	221	36	43.9	1442	2	C82898	DNA polymerase III
149	36	43.9	83	2	G82744	hypothetical prote	222	36	43.9	1643	2	T07961	myosin heavy chain
150	36	43.9	102	2	T31572	hypothetical prote	223	36	43.9	1707	2	A33526	collagen alpha 2(I
151	36	43.9	107	2	B97204	hypothetical prote	224	36	43.9	1759	2	T03725	replicase polyprot
152	36	43.9	118	1	B42959	14K hypothetical p	225	36	43.9	3519	2	S43048	polyketide synthas
153	36	43.9	118	2	H85889	hypothetical prote	226	36	43.9	3643	2	T36410	probable polyketid
154	36	43.9	118	2	E91045	hypothetical prote	227	35.5	43.3	110	2	T17913	hypothetical prote
155	36	43.9	118	2	T07905	low-carbon dioxide	228	35.5	43.3	320	2	S40405	MADS box protein o
156	36	43.9	133	2	D86175	hypothetical prote	229	35.5	43.3	327	2	AG3367	alcohol dehydrogen
157	36	43.9	143	2	S00718	ribosomal protein	230	35.5	43.3	384	2	G82976	probable rubredoxi
158	36	43.9	166	2	S28242	NADH2 dehydrogenas	231	35.5	43.3	472	2	AC3534	glu/asp-tRNA amido
159	36	43.9	197	1	RPBP16	repressor protein	232	35.5	43.3	546	2	T34480	serine-tRNA ligase
160	36	43.9	210	2	F72224	conserved hypothet	233	35.5	43.3	2347	1	TVHURS	kinase-related pro
161	36	43.9	223	2	F81611	ribosomal protein	234	35.5	43.3	2492	1	A44213	nonstructural poly
162	36	43.9	223	2	F86570	S3 ribosomal prote	235	35.5	43.3	3947	2	T52486	ferrichrome sidero
163	36	43.9	233	2	C72055	S3 ribosomal prote	236	35	42.7	88	2	AB3320	hypothetical prote
164	36	43.9	236	2	T02577	probable AP2 domai	237	35	42.7	94	2	A27691	methionine gamma-L
165	36	43.9	244	2	B86197	hypothetical prote	238	35	42.7	95	2	G37262	Ig heavy chain V r
166	36	43.9	247	2	T06721	hypothetical prote	239	35	42.7	104	2	E82360	cyat protein VC012
167	36	43.9	248	1	G64839	ymcB protein - Esc	240	35	42.7	107	2	G83348	hypothetical prote
168	36	43.9	248	2	E90771	hypothetical prote	241	35	42.7	138	2	S15441	ribosomal protein
169	36	43.9	248	2	A85634	hypothetical prote	242	35	42.7	138	2	AC3337	hypothetical prote
170	36	43.9	257	2	G84712	hypothetical prote	243	35	42.7	141	2	A70556	probable mutator M
171	36	43.9	260	2	S74597	ABC-type transport	244	35	42.7	147	2	AC1367	conserved hypothet
172	36	43.9	273	1	R5R22	ribosomal protein	245	35	42.7	147	2	AD1736	conserved hypothet
173	36	43.9	273	1	R5ZM2	ribosomal protein	246	35	42.7	153	2	S47943	translation initia
174	36	43.9	284	2	B87649	hypothetical prote	247	35	42.7	160	2	A10760	probable propanedi
175	36	43.9	298	2	C86871	conserved hypothet	248	35	42.7	183	2	S20844	modulation protein

249 35 42.7 183 2 G95313
250 35 42.7 186 1 AXPg
251 35 42.7 188 2 S66001
252 35 42.7 196 2 C84919
253 35 42.7 198 2 E83211
254 35 42.7 215 2 D81120
255 35 42.7 216 2 D97119
256 35 42.7 236 2 B84718
257 35 42.7 245 2 E86168
258 35 42.7 246 2 H95408
259 35 42.7 248 2 S03888
260 35 42.7 250 2 S41513
261 35 42.7 253 2 A83280
262 35 42.7 255 1 VGBE2E
263 35 42.7 257 2 B82837
264 35 42.7 258 2 JH0472
265 35 42.7 259 2 G85363
266 35 42.7 260 2 S10016
267 35 42.7 261 2 A80405
268 35 42.7 263 2 JC5271
269 35 42.7 266 2 A80665
270 35 42.7 268 2 AG3322
271 35 42.7 273 2 T47822
272 35 42.7 277 2 T03927
273 35 42.7 278 2 A75427
274 35 42.7 281 2 AF0505
275 35 42.7 290 2 JU0473
276 35 42.7 291 2 D70729
277 35 42.7 300 2 D70018
278 35 42.7 306 2 A10441
279 35 42.7 312 2 S53969
280 35 42.7 314 2 AC2868
281 35 42.7 314 2 G97644
282 35 42.7 317 2 F72246
283 35 42.7 317 2 AD3067
284 35 42.7 319 2 AH2062
285 35 42.7 321 2 S31088
286 35 42.7 323 2 T09550
287 35 42.7 326 2 T09548
288 35 42.7 327 1 S13827
289 35 42.7 327 2 A82241
290 35 42.7 329 2 G71140
291 35 42.7 329 2 T47795
292 35 42.7 338 2 F98219
293 35 42.7 338 2 B49642
294 35 42.7 341 2 F69171
295 35 42.7 349 2 E90453
296 35 42.7 352 2 F96548
297 35 42.7 370 2 A75371
298 35 42.7 370 2 B72594
299 35 42.7 382 2 AC3118
300 35 42.7 391 2 AG2536
301 35 42.7 398 2 JC4174
302 35 42.7 399 2 B71140
303 35 42.7 402 2 F98169
304 35 42.7 421 2 T33811
305 35 42.7 421 2 T42252
306 35 42.7 425 2 T12473
307 35 42.7 425 2 S14147
308 35 42.7 435 2 T20819
309 35 42.7 453 2 A48348
310 35 42.7 454 2 A82952
311 35 42.7 462 2 T17480
312 35 42.7 466 2 AF0190
313 35 42.7 469 1 NMIVH3
314 35 42.7 469 1 NMIVN2
315 35 42.7 486 1 COBYW2
316 35 42.7 488 2 B47032
317 35 42.7 488 2 A95301
318 35 42.7 498 1 VGXPLA
319 35 42.7 498 1 VGXPLC
320 35 42.7 498 1 VGXPLM
321 35 42.7 506 2 S75789

NodL Nod factor ac
adrenodoxin precu
conserved hypothet
hypothetical prote
hypothetical prote
probable periplasm
amidase from nicot
hypothetical prote
hypothetical prote
probable short cha
phosphatase II oxy
Brn-3c protein - m
hypothetical membr
glycoprotein E - e
conserved hypothet
apolipoprotein A-I
hypothetical prote
phosphatase II oxy
methionyl aminopep
oxygen-evolving co
conserved hypothet
anhydro-N-acetylmu
hypothetical prote
DNA binding protei
hypothetical prote
probable exported
ypua protein - bac
probable peptide t
sugar permease hom
probable membrane
hypothetical prote
UDP-hexose transfe
UDP-hexose transfe
glyoxalase - Ther
hypothetical prote
aldo/keto reductas
phosphoprotein pho
phosphoprotein pho
phosphoprotein pho
phosphoprotein pho
probable ABC trans
hypothetical prote
hypothetical prote
hypothetical prote
POU-domain protein
conserved hypothet
glycosyltransferas
probable DNA-bindi
threonine 3-dehydr
hypothetical prote
transcription regu
hypothetical prote
methionine gamma-1
methionine prote
hypothetical prote
frcR protein (AF19
hypothetical prote
polypeptide N-acet
hypothetical prote
multifunctional pu
hypothetical prote
genome polyprotein
glucosamine-1-phos
endo-xy lanase homo
probable transmemb
exo-alpha-sialidas
exo-alpha-sialidas
whf2 protein - yea
galactose-1-phosph
hypothetical prote
surface glycoprote
surface glycoprote
6-aminohexanoate-c

322 35 42.7 513 2 T46143
323 35 42.7 516 2 G70556
324 35 42.7 518 2 T05277
325 35 42.7 519 2 E87233
326 35 42.7 527 2 T27572
327 35 42.7 529 2 T45254
328 35 42.7 537 2 T45634
329 35 42.7 543 2 B82215
330 35 42.7 548 2 T42617
331 35 42.7 550 1 VGBEG5
332 35 42.7 552 1 A81236
333 35 42.7 558 2 AB1599
334 35 42.7 562 2 S25725
335 35 42.7 566 2 AGU158
336 35 42.7 578 2 B64012
337 35 42.7 581 2 JC7086
338 35 42.7 589 2 T12820
339 35 42.7 595 2 H81044
340 35 42.7 626 2 JQ2322
341 35 42.7 630 2 A47398
342 35 42.7 642 2 S55521
343 35 42.7 651 2 T12083
344 35 42.7 667 2 S56829
345 35 42.7 678 2 S77215
346 35 42.7 708 2 T19474
347 35 42.7 720 1 HJEC2
348 35 42.7 720 2 AG0918
349 35 42.7 720 2 G91221
350 35 42.7 720 2 B86068
351 35 42.7 720 2 F82482
352 35 42.7 722 2 E82353
353 35 42.7 723 2 D73607
354 35 42.7 747 2 T23607
355 35 42.7 761 2 D70750
356 35 42.7 781 2 AH3355
357 35 42.7 784 2 AH2560
358 35 42.7 795 2 T20609
359 35 42.7 841 2 A90669
360 35 42.7 841 2 D85519
361 35 42.7 841 2 C64755
362 35 42.7 910 2 A86782
363 35 42.7 925 2 T29585
364 35 42.7 962 2 D70661
365 35 42.7 968 2 C82452
366 35 42.7 994 2 AC3480
367 35 42.7 1046 2 T42734
368 35 42.7 1101 2 G70951
369 35 42.7 1194 2 E84499
370 35 42.7 1293 2 S24402
371 35 42.7 1460 1 RNBX3L
372 35 42.7 1530 2 AD1663
373 35 42.7 1582 2 AC1153
374 35 42.7 1728 2 T17466
375 35 42.7 4976 2 T14165
376 34.5 42.1 114 2 F0181
377 34.5 42.1 116 2 S09962
378 34.5 42.1 140 2 S09216
379 34.5 42.1 288 2 S18990
380 34.5 42.1 390 2 C70765
381 34.5 42.1 396 2 D82584
382 34.5 42.1 402 2 E70656
383 34.5 42.1 491 2 B96739
384 34.5 42.1 618 2 D88483
385 34 41.5 49 2 I48681
386 34 41.5 64 2 T07567
387 34 41.5 72 2 A26084
388 34 41.5 79 2 T18142
389 34 41.5 86 2 T17313
390 34 41.5 108 1 S77013
391 34 41.5 118 2 AF0816
392 34 41.5 129 1 MVECMT
393 34 41.5 132 2 G85492
394 34 41.5 132 2 G90641

steroid 22-alpha-h
probable trpE prot
dihydrofolate redu
probable secreted
hypothetical prote
probable anthranil
hypothetical prote
methyl-accepting c
probable envelope
glycoprotein E pre
glycoprotein E pre
glycerol 3 phospho
glycerol 3 phospho
poly(3-hydroxyalka
protein-Npi-phosph
hypothetical prote
F2D10 protein - hu
hypothetical prote
hemolysin activati
starch synthase (E
serotonin transpor
beta-fructofuranos
beta-fructofuranos
probable membrane
hypothetical prote
hypothetical prote
DNA helicase II (E
DNA helicase II (i
DNA-dependent ATPa
DNA-dependent ATPa
hypothetical prote
DNA helicase II VC
hypothetical prote
probable ctpA prot
outer membrane pro
hypothetical prote
hypothetical prote
probable enzyme (i
probable enzyme ya
yagK protein - Esc
cation-transporthin
cytoplasmic linker
probable ATP-depen
protein K04G7.3 (i
xeroderma pigmento
DNA-directed RNA p
glutamate synthase
adhesin homolog lm
rifamycin polyketi
peptide synthetase
Ig heavy chain V r
Ig heavy chain V-D
Ig heavy chain pre
probable acyl-CoA
probable cobL - My
penicillin binding
hypothetical prote
hypothetical prote
protein let-721 (i
185 kDa glycoposph
hypothetical prote
regulatory protein
hypothetical prote
hypothetical prote
hypothetical prote
conserved hypothet
7,8-dihydro-8-oxog
7,8-dihydro-8-oxog
7,8-dihydro-8-oxog

395	34	41.5	162	2	H75435	hypothetical prote	458	34	41.5	381	2	AF3303	putrescine transpo
396	34	41.5	163	2	T36335	probable inorganic	459	34	41.5	382	2	F87464	conserved hypotet
397	34	41.5	165	2	D69784	hypothetical prote	470	34	41.5	384	2	B86660	amino acid amidohy
398	34	41.5	168	2	AH1176	B. subtilis regula	471	34	41.5	384	2	JC5206	hypothetical 43.0K
399	34	41.5	168	2	D84732	hypothetical prote	472	34	41.5	389	2	B82819	outer membrane pro
400	34	41.5	171	2	T00432	probable AP2 domai	473	34	41.5	393	2	A82664	conserved hypotet
401	34	41.5	194	2	H84771	probable AP2 domai	474	34	41.5	401	2	T35956	probable acyl-CoA
402	34	41.5	202	2	S30164	strong gravity str	475	34	41.5	401	2	A97446	hypothetical prote
403	34	41.5	203	2	T04311	lexA protein - Pse	476	34	41.5	408	2	A82162	hypothetical prote
404	34	41.5	204	2	S30165	repressor protein	477	34	41.5	410	2	S38238	hypothetical prote
405	34	41.5	207	2	T49897	transcription fact	478	34	41.5	414	2	A69907	cell wall-binding
406	34	41.5	208	2	T17092	NADH2 dehydrogenas	479	34	41.5	429	2	C84194	hypothetical prote
407	34	41.5	216	2	C87254	hypothetical prote	480	34	41.5	432	2	D95369	conserved hypotet
408	34	41.5	224	2	D83570	probable nucleotid	481	34	41.5	434	2	G37610	hypothetical metab
409	34	41.5	225	2	A85196	EREBP-2 protein [i	482	34	41.5	434	2	AC2833	hypothetical prote
410	34	41.5	225	2	T52011	ethylene responsiv	483	34	41.5	457	2	T39751	major facilitator
411	34	41.5	233	2	T02590	DNA binding protei	484	34	41.5	459	2	A52390	thiophen / furan o
412	34	41.5	234	2	T07686	transcription fact	485	34	41.5	459	2	S28025	light harvesting c
413	34	41.5	236	2	T02432	ethylene-responsiv	486	34	41.5	469	1	NMIV2	exo-alpha-sialidas
414	34	41.5	237	1	S11927	licheninase (EC 3.	487	34	41.5	472	2	B75501	glycogen synthase
415	34	41.5	238	2	AB33442	sugar fermentation	488	34	41.5	478	2	G75052	pyruvate kinase (E
416	34	41.5	238	2	S70195	ktfA protein - Bsc	489	34	41.5	478	2	F71171	probable pyruvate
417	34	41.5	240	2	H82289	conserved hypotet	490	34	41.5	481	2	H96529	hypothetical prote
418	34	41.5	243	2	T51989	ethylene responsiv	491	34	41.5	483	2	JC7179	acid phosphatase (
419	34	41.5	245	2	F70583	hypothetical prote	492	34	41.5	488	2	T30914	xylan 1,4-beta-xy
420	34	41.5	254	2	A11117	dehydrogenase/redu	493	34	41.5	491	2	T49304	hypothetical prote
421	34	41.5	254	2	AC1478	dehydrogenase/redu	494	34	41.5	494	2	A58721	sulfate adenylyltr
422	34	41.5	254	2	F90107	60S ribosomal prot	495	34	41.5	498	2	S27849	variant surface gl
423	34	41.5	255	2	G97227	protein containing	496	34	41.5	502	2	C96960	uncharacterized co
424	34	41.5	256	1	BVECBH	biotin biosynthesi	497	34	41.5	503	2	S31940	starch-degrading e
425	34	41.5	256	2	E86006	biotin biosynthesi	498	34	41.5	503	2	C97262	N-terminal domain
426	34	41.5	256	2	F91160	biotin biosynthesi	499	34	41.5	529	2	AC2112	hypothetical prote
427	34	41.5	259	2	G84695	probable DOF zinc	500	34	41.5	547	2	T00977	probable pectinest
428	34	41.5	259	2	C69122	hypothetical prote	501	34	41.5	554	2	H86772	alpha-acetolactate
429	34	41.5	261	2	B69095	cobalt transport m	502	34	41.5	563	1	A47078	phenol catabolic p
430	34	41.5	261	2	G84376	hypothetical prote	503	34	41.5	563	1	S47095	phnK protein - Pse
431	34	41.5	262	2	E96747	hypothetical prote	504	34	41.5	563	1	D84717	probable cysteinyl
432	34	41.5	262	2	S29396	ribosomal protein	505	34	41.5	566	1	NIPSRP	regulatory protein
433	34	41.5	266	1	R5NT2D	ethylene responsiv	506	34	41.5	574	2	S23530	H+-transporting tw
434	34	41.5	266	2	T51988	hypothetical prote	507	34	41.5	574	2	AF1108	transporter homolo
435	34	41.5	268	2	S05471	embryonic abundant	508	34	41.5	628	2	AG1469	transporter homolo
436	34	41.5	274	1	R5NT2	ribosomal protein	509	34	41.5	628	2	AF1108	transporter homolo
437	34	41.5	274	2	S78397	ribosomal protein	510	34	41.5	635	2	S76371	hypothetical prote
438	34	41.5	286	1	R5SP2	ribosomal protein	511	34	41.5	635	2	F90561	hypothetical prote
439	34	41.5	289	2	AC0794	conserved hypotet	512	34	41.5	673	1	CGB06C	collagen alpha 1(I
440	34	41.5	301	2	AC0295	conserved hypotet	513	34	41.5	674	2	T05264	probable serine/th
441	34	41.5	301	2	S77169	hypothetical prote	514	34	41.5	681	2	H83044	probable fuaA prot
442	34	41.5	302	1	PAFF1A	phosphoprotein pho	515	34	41.5	701	2	E70827	hypothetical prote
443	34	41.5	302	2	S29396	phosphoprotein pho	516	34	41.5	712	2	T48961	hypothetical prote
444	34	41.5	302	2	AE2520	hypothetical prote	517	34	41.5	729	2	F83725	alcohol dehydrogen
445	34	41.5	303	2	B43451	Na+/K+-exchanging	518	34	41.5	742	2	A90326	hypothetical prote
446	34	41.5	305	2	S44767	C99E4.1 protein -	519	34	41.5	749	2	A90568	hypothetical prote
447	34	41.5	311	2	T00087	riboseyltransfera	520	34	41.5	751	2	AD2168	5-methylterahydro
448	34	41.5	312	2	C87562	conserved hypotet	521	34	41.5	751	2	B97936	probable transcrip
449	34	41.5	313	2	AC0622	probable bacteriop	522	34	41.5	857	2	T40824	beta-N-acetylgluco
450	34	41.5	313	2	AC0734	probable bacteriop	523	34	41.5	863	2	JC7537	granulocyte colony
451	34	41.5	316	2	S25188	transcription acti	524	34	41.5	863	2	C38252	penicillin-binding
452	34	41.5	321	2	E75540	hypothetical prote	525	34	41.5	864	2	JH0438	hypothetical prote
453	34	41.5	335	2	H86363	F19G10.7 protein -	526	34	41.5	903	2	T44153	H+-exporting ATPas
454	34	41.5	337	2	G83200	probable oxidoredu	527	34	41.5	905	2	S75035	Magnesium transport
455	34	41.5	340	2	S11143	class I histocoppa	528	34	41.5	908	2	AC0967	Mg2+-transporting
456	34	41.5	341	2	AI1039	phage integrase [i	529	34	41.5	908	2	B39083	GGDEF family prote
457	34	41.5	342	2	QJ1651	SSL2 protein - hum	530	34	41.5	976	2	G82209	protein F40F12.5 [
458	34	41.5	344	2	T39023	hypothetical prote	531	34	41.5	1021	2	F88568	probable limonene
459	34	41.5	345	2	C71046	hypothetical prote	532	34	41.5	1024	2	G71434	DNA polymerase I (
460	34	41.5	349	2	A75169	probable iron (III	533	34	41.5	1031	2	B97372	DNA polymerase I (
461	34	41.5	352	2	D81718	tRNA (Guanine-N1) -	534	34	41.5	1031	2	A12589	probable membrane
462	34	41.5	358	2	T52073	ER-associated Hsp4	535	34	41.5	1033	2	S54506	hypothetical prote
463	34	41.5	365	2	S10847	collagen alpha 2(I	536	34	41.5	1050	2	T27753	hypothetical prote
464	34	41.5	374	2	T46065	hypothetical prote	537	34	41.5	1099	2	T18713	hypothetical prote
465	34	41.5	374	2	T43966	hypothetical prote	538	34	41.5	1120	2	T14275	myosin-like protei
466	34	41.5	376	2	S45763	hypothetical prote	539	34	41.5	1130	2	A48843	MHC class II trans
467	34	41.5	377	2	AE0589	galactosyltransfer	540	34	41.5	1153	2	A97179	ATP-dependent exon
										1191	2	S27329	DNA topoisomerase

541	34	41.5	1192	1	ISXFA5	DNA topoisomerase	614	33	40.2	242	2	G83948	3-oxoacyl-[acyl ca
542	34	41.5	1194	2	E96624	hypothetical prote	615	33	40.2	244	2	A96806	hypothetical prote
543	34	41.5	1263	2	T19472	hypothetical prote	616	33	40.2	246	2	AC2504	hypothetical prote
544	34	41.5	1337	2	T30291	dextranase - Strept	617	33	40.2	250	2	T01604	hypothetical prote
545	34	41.5	1392	2	A43336	microtubule-vesicl	618	33	40.2	254	2	D95360	probable GntR-fami
546	34	41.5	1418	2	T45467	collagen alpha 1(I)	619	33	40.2	261	2	T04301	beta-expansin - ri
547	34	41.5	1419	2	A41182	collagen alpha 1(I)	620	33	40.2	262	2	S49311	2,4-dihydroxyhept-
548	34	41.5	1427	2	S74293	SRB8 protein - yea	621	33	40.2	262	2	T04301	hypothetical prote
549	34	41.5	1427	2	S22695	restin - human	622	33	40.2	263	2	A95334	hypothetical prote
550	34	41.5	1486	1	B40333	collagen alpha 1(I)	623	33	40.2	268	2	AD2517	hypothetical prote
551	34	41.5	1487	1	CGHUC6	collagen alpha 1(I)	624	33	40.2	271	2	S14068	seed protein precu
552	34	41.5	1487	2	B41182	collagen alpha 1(I)	625	33	40.2	275	2	T11810	aminoglycoside N3'
553	34	41.5	1492	2	A40333	collagen alpha 1(I)	626	33	40.2	275	2	T02334	ribosomal protein
554	34	41.5	1763	2	T17465	collagen alpha 1(I)	627	33	40.2	275	2	T52333	probable urease ac
555	34	41.5	1846	2	T10670	rifamycin polyketi	628	33	40.2	275	2	G90221	conserved hypotet
556	34	41.5	2254	2	T09053	low voltage-activa	629	33	40.2	275	2	A84429	probable S-lucos g
557	34	41.5	4116	2	T13719	calo protein - fru	630	33	40.2	276	2	T07531	ribosomal protein
558	34	41.5	4735	2	T17463	rifamycin polyketi	631	33	40.2	276	2	T47351	hypothetical prote
559	34	41.5	6420	2	T30283	polyketide synthas	632	33	40.2	280	2	T00718	calcium-dependent
560	34	41.5	26926	1	T38344	titin, cardiac mus	633	33	40.2	288	2	D85072	hypothetical prote
561	33.5	40.9	59	2	F69854	hypothetical prote	634	33	40.2	291	2	T20666	hypothetical prote
562	33.5	40.9	123	2	S30529	Ig heavy chain V r	635	33	40.2	292	2	E70018	sugar permease hom
563	33.5	40.9	124	2	B55257	Ig gamma heavy cha	636	33	40.2	293	1	DBP32	helix-destabilizin
564	33.5	40.9	205	2	H70468	phosphoribosyl-AMP	637	33	40.2	293	1	DBP32	helix-destabilizin
565	33.5	40.9	239	2	C70486	conserved hypotet	638	33	40.2	297	2	AF0568	carbamate kinase (
566	33.5	40.9	254	2	S34724	probable oxidoredu	639	33	40.2	299	2	F83211	hypothetical prote
567	33.5	40.9	284	2	T09843	amino acid transpo	640	33	40.2	299	2	A56663	capsid protein VPI
568	33.5	40.9	289	2	G97122	pseudouridine synt	641	33	40.2	300	2	T33232	hypothetical prote
569	33.5	40.9	351	2	T18066	hypothetical prote	642	33	40.2	301	1	DBP34	helix-destabilizin
570	33.5	40.9	381	2	E71194	probable N2,N2-dim	643	33	40.2	301	2	H95308	probable ABC trans
571	33.5	40.9	422	2	G72227	hypothetical prote	644	33	40.2	308	2	S11153	oligopeptide trans
572	33.5	40.9	486	2	T10100	amino acid transpo	645	33	40.2	308	2	E95220	hypothetical prote
573	33.5	40.9	503	2	E91055	hypothetical prote	646	33	40.2	308	2	T35303	hypothetical prote
574	33.5	40.9	503	2	A85900	hypothetical prote	647	33	40.2	311	2	B95222	hypothetical prote
575	33.5	40.9	954	2	S46105	glucan 1,4-alpha-g	648	33	40.2	311	2	A99086	bioi-facetyl-CoA-
576	33.5	40.9	1112	2	S70522	cyclic nucleotide	649	33	40.2	314	2	AH2517	hypothetical prote
577	33.5	40.9	1353	2	T00347	hypothetical prote	650	33	40.2	316	2	T27194	hypothetical prote
578	33.5	40.9	1401	2	T30247	Werner syndrome pr	651	33	40.2	317	2	D87395	metallo-beta-lacta
579	33	40.2	20	2	PM0003	chlorophyll a/b-bi	652	33	40.2	323	2	E98084	hypothetical prote
580	33	40.2	61	2	B34123	depressant insect	653	33	40.2	323	2	B48067	ethanolamine-phosp
581	33	40.2	84	2	E83004	glutaredoxin PA512	654	33	40.2	325	1	A71887	probable GMP reduc
582	33	40.2	89	2	A97705	hypothetical prote	655	33	40.2	325	2	T26180	hypothetical prote
583	33	40.2	91	2	T05920	probable cysteine	656	33	40.2	325	2	T29604	hypothetical prote
584	33	40.2	97	2	AB3591	hypothetical prote	657	33	40.2	327	1	F64626	probable GMP reduc
585	33	40.2	102	2	T17738	hypothetical prote	658	33	40.2	328	2	T44931	moCR protein limpo
586	33	40.2	103	2	T10920	3C3.11 protein - S	659	33	40.2	329	2	F72413	conserved hypotet
587	33	40.2	105	2	A13472	hypothetical prote	660	33	40.2	330	2	T51834	transcription fact
588	33	40.2	115	2	E97483	hypothetical prote	661	33	40.2	332	1	WMBP16	gene 16 protein -
589	33	40.2	122	2	T44906	hypothetical prote	662	33	40.2	332	1	WMBP26	gene 16 protein -
590	33	40.2	126	2	A23034	hypothetical prote	663	33	40.2	332	2	B75011	hypothetical prote
591	33	40.2	127	2	A54670	RNA polymerase II	664	33	40.2	342	2	F85089	arginase [imported
592	33	40.2	128	2	E72804	gp39 protein - Myc	665	33	40.2	343	2	T36590	hypothetical prote
593	33	40.2	133	2	G64445	formylmethanofuran	666	33	40.2	343	2	AB1955	hypothetical prote
594	33	40.2	135	2	G87686	hypothetical prote	667	33	40.2	352	2	F82284	S-adenosylmethioni
595	33	40.2	136	2	T34316	hypothetical prote	668	33	40.2	353	2	C85572	hypothetical prote
596	33	40.2	139	2	S76176	hypothetical prote	669	33	40.2	353	2	E90721	hypothetical prote
597	33	40.2	157	2	B72753	hypothetical prote	670	33	40.2	354	2	C82038	nitrogen regulatio
598	33	40.2	157	2	D84053	hypothetical prote	671	33	40.2	355	2	F70983	probable serine pr
599	33	40.2	159	2	AG3609	transcription regu	672	33	40.2	361	2	A86386	probable DNA-bindi
600	33	40.2	160	2	JO0542	185K secretory pro	673	33	40.2	362	2	C64807	ybgO protein - Esc
601	33	40.2	171	2	D55853	aggA protein precu	674	33	40.2	362	2	T18782	hypothetical prote
602	33	40.2	185	2	D69976	conserved hypotet	675	33	40.2	362	2	T42689	hypothetical prote
603	33	40.2	196	2	A48150	hibernation-relate	676	33	40.2	367	2	H64696	lipopolysaccharide
604	33	40.2	203	2	G86785	acetyltransferase	677	33	40.2	367	2	H71822	lipopolysaccharide
605	33	40.2	203	2	H86834	maltose O-acetyltr	678	33	40.2	369	2	B70968	hypothetical prote
606	33	40.2	203	2	S53587	probable membrane	679	33	40.2	369	2	F95353	probable ABC trans
607	33	40.2	207	2	T17084	NADH2 dehydrogenas	680	33	40.2	373	2	T26030	hypothetical prote
608	33	40.2	208	2	T17087	NADH2 dehydrogenas	681	33	40.2	377	2	C69858	conserved hypotet
609	33	40.2	211	2	D96507	hypothetical prote	682	33	40.2	379	2	G87046	probable acyltrans
610	33	40.2	226	2	H81295	probable two-compo	683	33	40.2	384	2	G75567	conserved hypotet
611	33	40.2	230	2	D86352	protein T26F17.14	684	33	40.2	386	1	A29984	alanine racemase (
612	33	40.2	239	2	E81697	ribose 5-phosphate	685	33	40.2	386	2	T44545	contractile tail s
613	33	40.2	241	2	A53014	chloride conductan	686	33	40.2	387	2	H72299	hypothetical prote

687	33	40.2	387	2	H72273	conserved hypothet	760	33	40.2	628	2	T09785	hydroxymethylgluta
688	33	40.2	389	2	G95939	probable hippurath	761	33	40.2	634	2	S31925	beta-fructofuranos
689	33	40.2	392	2	AE2016	phosphoserine amin	762	33	40.2	636	1	S31157	beta-fructofuranos
690	33	40.2	394	2	S62009	probable membrane	763	33	40.2	639	1	S31155	beta-fructofuranos
691	33	40.2	394	2	AD0842	probable transmembr	764	33	40.2	639	2	A56126	peroxisomal target
692	33	40.2	395	2	T33677	hypothetical prote	765	33	40.2	640	1	A86657	fructose-bisphosph
693	33	40.2	398	2	A95870	hypothetical prote	766	33	40.2	640	2	T09534	probable beta-fruc
694	33	40.2	399	2	D83535	aromatic amino aci	767	33	40.2	640	2	AB2251	glucose inhibited
695	33	40.2	403	2	A10699	probable pathogeni	768	33	40.2	644	2	D83971	stage V sporulatio
696	33	40.2	404	1	S62440	mevalonate kinase	769	33	40.2	656	2	G85731	Rhs element associ
697	33	40.2	408	2	B71047	probable ferredoxi	770	33	40.2	661	2	S37591	beta-fructofuranos
698	33	40.2	408	2	C86903	hypothetical prote	771	33	40.2	666	2	T08904	probable long-chai
699	33	40.2	408	2	AC1840	hypothetical prote	772	33	40.2	667	2	T07929	probable long-chai
700	33	40.2	409	2	T35598	hypothetical prote	773	33	40.2	681	2	D84019	methylmalonyl-CoA
701	33	40.2	416	2	D83386	hypothetical prote	774	33	40.2	686	2	JC5708	villin-like protei
702	33	40.2	420	1	DCECD	diaminopimelate de	775	33	40.2	694	2	C72761	hypothetical prote
703	33	40.2	420	2	AG0867	diaminopimelate de	776	33	40.2	698	2	E85369	hypothetical prote
704	33	40.2	420	2	B85936	diaminopimelate de	777	33	40.2	698	2	T10682	hypothetical prote
705	33	40.2	420	2	G91090	diaminopimelate de	778	33	40.2	698	2	AD2985	oxidoreductase Atu
706	33	40.2	429	2	C75013	phosphoribosylglyc	779	33	40.2	706	2	F87683	peptidase M13 fami
707	33	40.2	433	2	C71138	phosphoribosylglyc	780	33	40.2	708	2	S52317	quinohemoprotein e
708	33	40.2	437	2	S73284	hypothetical prote	781	33	40.2	714	2	T16126	hypothetical prote
709	33	40.2	437	2	G82777	glutamate symport	782	33	40.2	728	2	H82965	DNA helicase II PA
710	33	40.2	439	2	S16530	xylose isomerase (783	33	40.2	729	2	C98298	probable oxidoredu
711	33	40.2	445	2	G81200	UDP-N-acetylmutamo	784	33	40.2	735	2	A69146	hypothetical prote
712	33	40.2	445	2	D81777	UDP-N-acetylmutamo	785	33	40.2	737	2	AE1306	heavy metal-transp
713	33	40.2	452	2	T28094	hypothetical prote	786	33	40.2	737	2	AE1678	heavy metal-transp
714	33	40.2	455	2	T23712	hypothetical prote	787	33	40.2	737	2	C71122	hypothetical prote
715	33	40.2	458	2	A90565	hypothetical prote	788	33	40.2	737	2	S76989	hypothetical prote
716	33	40.2	464	2	T28818	hypothetical prote	789	33	40.2	775	2	A61228	sensory transducti
717	33	40.2	466	2	AH3100	amidohydrolase lim	790	33	40.2	786	2	A35466	collagen alpha 2(I
718	33	40.2	466	2	B98186	probable hydrolase	791	33	40.2	797	2	B95377	progesterone recep
719	33	40.2	467	2	I56896	gene gfi-2 protein	792	33	40.2	807	2	T24110	hypothetical prote
720	33	40.2	469	2	J01644	exo-alpha-sialidas	793	33	40.2	808	1	OPKEX	hypothetical prote
721	33	40.2	471	2	S08333	exo-alpha-sialidas	794	33	40.2	819	2	A47018	glucose dehydrogen
722	33	40.2	471	2	T50016	transcription fact	795	33	40.2	819	2	B87386	lectin-like adhesi
723	33	40.2	474	2	T03126	hypothetical prote	796	33	40.2	835	2	F80076	hypothetical prote
724	33	40.2	474	2	E90318	medium-chain-fatty	797	33	40.2	843	2	A27131	epidermal growth f
725	33	40.2	475	2	E86249	hypothetical prote	798	33	40.2	872	2	S73948	Mg2+-transporting
726	33	40.2	477	2	S34547	H+-transporting tw	799	33	40.2	878	2	B84977	alanine-tRNA ligas
727	33	40.2	485	2	A84859	probable cytochrom	800	33	40.2	899	2	AG0202	Mg(2+) transport A
728	33	40.2	489	2	S63401	hypothetical prote	801	33	40.2	942	2	C96574	hypothetical prote
729	33	40.2	490	2	E83202	outer membrane pro	802	33	40.2	944	2	AC2073	two-component sens
730	33	40.2	490	2	JS0586	algG protein precu	803	33	40.2	959	2	T00246	DNA polymerase V -
731	33	40.2	490	2	T41039	probable transcrip	804	33	40.2	974	2	T14076	probable villin [i
732	33	40.2	492	2	AH2809	hypothetical prote	805	33	40.2	984	2	AE0290	insecticidal toxin
733	33	40.2	495	2	B35721	nicotinic acetylch	806	33	40.2	993	2	A38437	probable homeotic
734	33	40.2	499	2	AD2262	amidophosphoribosy	807	33	40.2	1008	2	H72310	conserved hypothet
735	33	40.2	507	2	B83988	proline transport	808	33	40.2	1017	2	T08553	hypothetical prote
736	33	40.2	509	2	T06226	probable beta-fruc	809	33	40.2	1019	2	T11560	pol polyprotein -
737	33	40.2	509	2	H87389	conserved hypothet	810	33	40.2	1025	1	JC1266	beta-galactosidase
738	33	40.2	516	2	AH2417	hypothetical prote	811	33	40.2	1029	2	T05050	protein kinase hom
739	33	40.2	517	2	A30992	probable nicotinic	812	33	40.2	1036	1	GNLJG2	HIV-1 retropepsin
740	33	40.2	520	2	S45702	leukotriene-B4 20-	813	33	40.2	1039	2	S46347	pol polyprotein -
741	33	40.2	520	2	S45753	probable membrane	814	33	40.2	1046	2	T42720	cytoplasmic linker
742	33	40.2	523	2	C97588	ribose ABC transpo	815	33	40.2	1050	2	G86582	exodeoxyribonuclea
743	33	40.2	536	2	A36395	spore wall maturat	816	33	40.2	1050	2	H72041	exodeoxyribonuclea
744	33	40.2	544	2	T07593	pectinesterase (EC	817	33	40.2	1054	2	C81624	HIV-1 retropepsin
745	33	40.2	548	2	C82698	electron transfer	818	33	40.2	1054	1	GNLJG5	HIV-1 retropepsin
746	33	40.2	555	2	D90369	NADH oxidase SSO20	819	33	40.2	1056	1	GNLJG3	pol polyprotein -
747	33	40.2	555	2	C83444	probable AMP-bindi	820	33	40.2	1058	2	S08436	hypothetical prote
748	33	40.2	556	2	A44441	B-cell antigen CD1	821	33	40.2	1070	2	T34385	replication factor
749	33	40.2	557	4	EBRTMS	IGF-binding protei	822	33	40.2	1092	2	T18305	replication factor
750	33	40.2	563	1	S11175	choline transport	823	33	40.2	1092	2	T18306	probable histidine
751	33	40.2	569	2	A46462	r cell activation	824	33	40.2	1122	2	T00441	DNA-directed DNA p
752	33	40.2	575	2	C83313	probable type II s	825	33	40.2	1139	2	D64503	protein R06B10.1 l
753	33	40.2	576	2	T40476	hypothetical prote	826	33	40.2	1236	2	A88392	cryptic nitrate re
754	33	40.2	586	2	B90659	Rhs core protein [827	33	40.2	1246	2	G90887	cryptic nitrate re
755	33	40.2	586	2	H85509	hypothetical prote	828	33	40.2	1246	2	B85730	nitrate reductase
756	33	40.2	600	2	H70448	G-protein LepA - A	829	33	40.2	1246	2	G84899	respiratory nitrat
757	33	40.2	607	2	A95122	TnS252, relaxase [830	33	40.2	1247	2	AH0648	hypothetical prote
758	33	40.2	608	2	T06632	hypothetical prote	831	33	40.2	1262	2	T25168	epidermal growth f
759	33	40.2	612	2	T08602	protein TipD - sli	832	33	40.2	1330	1	G0FFE	

833	33	40.2	1374	2	D85390	myosin-like protei	906	32	39.0	96	2	B41979	neuropeptide Y pre
834	33	40.2	1375	2	T05200	myosin heavy chain	907	32	39.0	98	2	PH1147	Ig heavy chain V r
835	33	40.2	T13822	2	T13822	frazzled gene prot	908	32	39.0	99	2	D38601	Ig kappa chain V r
836	33	40.2	T29637	2	A29637	position-specific	909	32	39.0	100	2	B70977	hypothetical prote
837	33	40.2	T29637	2	A29637	RhaD core protein	910	32	39.0	101	1	GMHUB	gastirin precursor
838	33	40.2	T29637	2	A29637	RhaD core protein	911	32	39.0	102	2	A82456	probable acetyltra
839	33	40.2	T29637	2	A29637	RhaG core protein	912	32	39.0	103	2	B81964	conserved hypoteth
840	33	40.2	T29637	2	A29637	RhaG core protein	913	32	39.0	104	2	C81020	cyay protein NMA19
841	33	40.2	T29637	2	A29637	RhaG core protein	914	32	39.0	105	2	A27563	Ig heavy chain V r
842	33	40.2	T29637	2	A29637	RhaD core protein	915	32	39.0	106	2	T03081	hypothetical prote
843	33	40.2	T29637	2	A29637	RhaD core protein	916	32	39.0	107	2	B85573	succinate dehydrog
844	33	40.2	T29637	2	A29637	probable ATP-depen	917	32	39.0	108	2	C90722	succinate dehydrog
845	33	40.2	T29637	2	A29637	frazzled gene prot	918	32	39.0	109	2	D86510	hypothetical prote
846	33	40.2	T29637	2	A29637	glutamate synthase	919	32	39.0	110	2	B72111	hypothetical prote
847	33	40.2	T29637	2	A29637	hypothetical prote	920	32	39.0	111	2	H90664	hypothetical prote
848	33	40.2	T29637	2	A29637	peptidoglycan boun	921	32	39.0	112	2	G95319	hypothetical prote
849	33	40.2	T29637	2	A29637	hypothetical glyci	922	32	39.0	113	2	C85515	unknown protein en
850	33	40.2	T29637	2	A29637	Balbani ring 3 pr	923	32	39.0	114	2	AC0264	probable pyrophosp
851	33	40.2	T29637	2	A29637	collagen alpha 2(I	924	32	39.0	115	2	F69044	mutator MutT prote
852	33	40.2	T29637	2	A29637	cell surface glyco	925	32	39.0	116	2	F86816	hypothetical prote
853	33	40.2	T29637	2	A29637	hypothetical prote	926	32	39.0	117	2	F87352	mutator mutT prote
854	33	40.2	T29637	2	A29637	myosin-light-chain	927	32	39.0	118	2	C86521	hypothetical prote
855	33	40.2	T29637	2	A29637	calcium channel pr	928	32	39.0	119	2	G72100	hypothetical prote
856	33	40.2	T29637	2	A29637	myosin heavy chain	929	32	39.0	120	2	C82466	conserved hypoteth
857	33	40.2	T29637	2	A29637	calcium channel BI	930	32	39.0	121	1	R3IT8	ribosomal protein
858	33	40.2	T29637	2	A29637	hypothetical prote	931	32	39.0	122	2	T00763	hypothetical prote
859	33	40.2	T29637	2	A29637	hypothetical prote	932	32	39.0	123	2	AC0864	conserved hypoteth
860	33	40.2	T29637	2	A29637	calcium channel BI	933	32	39.0	124	1	A55692	anaerobic ribonucl
861	33	40.2	T29637	2	A29637	hypothetical prote	934	32	39.0	125	2	AB1058	anaerobic ribonucl
862	33	40.2	T29637	2	A29637	cyclic beta 1-2 gl	935	32	39.0	126	2	F91280	anaerobic ribonucl
863	33	40.2	T29637	2	A29637	variant-specific s	936	32	39.0	127	2	F86121	hypothetical prote
864	33	40.2	T29637	2	A29637	hypothetical prote	937	32	39.0	128	2	H83160	hypothetical prote
865	33	40.2	T29637	2	A29637	hypothetical prote	938	32	39.0	129	2	T27757	hypothetical prote
866	33	40.2	T29637	2	A29637	rifamycin polyketi	939	32	39.0	130	2	AD3506	sensor protein chv
867	33	40.2	T29637	2	A29637	polyketide synthas	940	32	39.0	131	2	B71046	hypothetical prote
868	32.5	39.6	101	2	S13692	Ig heavy chain V r	941	32	39.0	132	2	F86186	YUP812.17 import
869	32.5	39.6	111	2	S13693	Ig heavy chain V r	942	32	39.0	133	2	B90398	mutT-like protein
870	32.5	39.6	112	2	S13694	Ig heavy chain V r	943	32	39.0	134	2	H90412	hypothetical prote
871	32.5	39.6	113	2	S13695	Ig heavy chain V r	944	32	39.0	135	2	T48744	hypothetical prote
872	32.5	39.6	114	2	S13696	Ig heavy chain V r	945	32	39.0	136	2	C75295	MutT/nudix family
873	32.5	39.6	115	2	S13697	Ig heavy chain V r	946	32	39.0	137	2	C83269	hypothetical prote
874	32.5	39.6	116	2	S13698	Ig heavy chain V r	947	32	39.0	138	2	S67663	hypothetical prote
875	32.5	39.6	117	2	S13699	Ig heavy chain V r	948	32	39.0	139	2	S36749	transcription fact
876	32.5	39.6	118	2	S13700	Ig heavy chain V r	949	32	39.0	140	2	AH3251	conserved hypoteth
877	32.5	39.6	119	2	S13701	Ig heavy chain V r	950	32	39.0	141	2	G87307	hypothetical prote
878	32.5	39.6	120	2	T49852	hypothetical prote	951	32	39.0	142	1	A42840	ribosomal protein
879	32.5	39.6	121	2	D75428	hypothetical prote	952	32	39.0	143	2	JQ0345	ubiquinol-cytochro
880	32.5	39.6	122	2	B71514	hypothetical prote	953	32	39.0	144	2	T49508	hypothetical prote
881	32.5	39.6	123	2	A57504	ycli protein impor	954	32	39.0	145	2	C81557	KDO-transferrase 2
882	32.5	39.6	124	2	S06824	conserved hypoteth	955	32	39.0	146	2	S12205	hypothetical prote
883	32.5	39.6	125	2	A84953	pantoate-beta-alan	956	32	39.0	147	2	A13391	hypothetical cytos
884	32.5	39.6	126	2	D75286	serine proteinase	957	32	39.0	148	2	T35037	hypothetical prote
885	32.5	39.6	127	2	H83253	conserved hypoteth	958	32	39.0	149	2	D90526	conserved hypoteth
886	32.5	39.6	128	2	S46993	elk ligand - human	959	32	39.0	150	2	JC7239	peroxiredoxin V -
887	32.5	39.6	129	2	T17584	probable beta-1,3-	960	32	39.0	151	2	JE0297	transcription fact
888	32.5	39.6	130	2	D64366	hypothetical prote	961	32	39.0	152	2	E82679	chaperone Xf1452 [
889	32.5	39.6	131	2	C69222	hypothetical prote	962	32	39.0	153	2	JE0297	DRE/CRT-binding pr
890	32.5	39.6	132	2	JC1497	alpha-amino-epsilo	963	32	39.0	154	2	JE0299	transcription facto
891	32.5	39.6	133	2	H81228	conserved hypoteth	964	32	39.0	155	2	T51830	transcription facto
892	32.5	39.6	134	2	C82000	probable integral	965	32	39.0	156	2	F70073	hypothetical prote
893	32.5	39.6	135	2	T12495	hypothetical prote	966	32	39.0	157	2	AC0084	probable DedA-fami
894	32.5	39.6	136	2	B82994	probable secretion	967	32	39.0	158	2	T00409	ethylene-responsiv
895	32.5	39.6	137	2	A05269	collagen alpha 1(I	968	32	39.0	159	2	T21380	hypothetical prote
896	32.5	39.6	138	2	T18892	hypothetical prote	969	32	39.0	160	2	T05800	probable transcrip
897	32.5	39.6	139	2	S62481	hypothetical prote	970	32	39.0	161	2	A38360	heat shock protein
898	32.5	39.6	140	2	T00056	hypothetical prote	971	32	39.0	162	2	E87550	hypothetical prote
899	32.5	39.6	141	2	T05663	hypothetical prote	972	32	39.0	163	2	S43513	hemoglobin linker
900	32.5	39.6	142	2	A13431	chromosome segrega	973	32	39.0	164	2	G75261	cytidylate kinase
901	32	39.0	143	2	T02859	probable serine/th	974	32	39.0	165	2	B83709	hypothetical prote
902	32	39.0	144	2	S71917	hemoglobin, extrac	975	32	39.0	166	2	E89899	uridylylate kinase
903	32	39.0	145	2	T08334	hypothetical prote	976	32	39.0	167	2	B71128	probable iron (III
904	32	39.0	146	2	S38920	hypothetical prote	977	32	39.0	168	2	T02511	DREB-like AP2 doma
905	32	39.0	147	2	G70611	hypothetical prote	978	32	39.0	169	2	JC2379	cell-specific heli


```

C;Genetics:
A;Gene: ECa3677
C;Superfamily: exodeoxyribonuclease V 135K chain

Query Match      56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 KKVWLGETSSAY 15
Db      1082 KSNWLGEDSSAY 1093

RESULT 5
AG2501
hypothetical protein all1791 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AG2501
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2501
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1596 <KUR>
A;Cross-references: UNIPROT:Q8YKV0; UNIPARC:UPI00000CEFI7; GB:BA000020; PIDN:BA078275.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1791
A;Genome: plasmid

Query Match      55.5%; Score 45.5; DB 2; Length 1596;
Best Local Similarity 52.9%; Pred. No. 35;
Matches 9; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY      1 RPGLYKQIWLGETSSA 14
      ||| :||| :
Db      240 RPGLYKQIWLGETSSA 256

RESULT 6
D84828
AP2 domain transcription factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: D84828
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84828
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: UNIPROT:Q9SIZ0; UNIPARC:UPI00000A8651; GB:AE002093; NID:G4586052; PI
C;Genetics:
A;Gene: At2g40350
A;Map position: 2

Query Match      54.9%; Score 45; DB 2; Length 177;
Best Local Similarity 61.5%; Pred. No. 4.9;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      3 GKKVWLGETSSAY 15
      ||| :||| :
Db      88 GAKLWLGETSSAY 100

C;Genetics:
A;Gene: ECa3677
C;Superfamily: exodeoxyribonuclease V 135K chain

Query Match      56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 KKVWLGETSSAY 15
Db      1082 KSNWLGEDSSAY 1093

RESULT 7
AD0783
PTS system, fructose-specific IIBC component [imported] - Salmonella enterica subsp. ent
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0783
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-562 <PAR>
A;Cross-references: UNIPARC:UPI0000059BC9; GB:AL513382; PIDN:CAD02586.1; PID:G16503443; C
C;Genetics:
A;Gene: STY2439
C;Superfamily: phosphotransferase system enzyme II, fructose-specific; phosphotransferase
Query Match      54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      3 GKKVWLGETSSA 14
      ||| :||| :
Db      59 GKKVWLGETSSA 70

RESULT 8
E84594
AP2 domain transcription factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: E84594
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <STO>
A;Cross-references: UNIPROT:Q9SKT1; UNIPARC:UPI000009FA1A; GB:AE002093; NID:G4454460; PI
C;Genetics:
A;Gene: At2g20880
A;Map position: 2

Query Match      53.7%; Score 44; DB 2; Length 336;
Best Local Similarity 58.3%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 PGKKVWLGETSS 13
      ||| :||| :
Db      43 PGSDMWLGDASS 54

RESULT 9
E85855
PTS system, fructose-specific transport protein [imported] - Escherichia coli (strain O15
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: E85855
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

```


| ||||| :||
Db 1082 KSNWLGSAAAY 1093

RESULT 16
C84828
AP2 domain transcription factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: C84828
R.;Lin, X.J.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Nishida, K.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.M.; Koo, H.; Moffat, K.S.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.; Eickholt, D.; Niernman, W.C.; Olsen, M.V.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84828
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-283 <STO>
A;Cross-references: UNIPROT:Q8LFR2; UNIPARC:UPI00000A3241; GB:AE002093; NID:g4586051; PIR:S00000
C;Genetics:
A;Gene: At2g40340
A;Map position: 2

Query Match 51.2%; Score 42; DB 2; Length 283;
Best Local Similarity 53.8%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKXVWLGTSAY 15
| :||| |::|
DB 35 GARLWLGTFSSSY 47

RESULT 17
C83034
Probable oxidoreductase PA4889 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 05-Oct-2004
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brinkman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
adman, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83034
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <STO>
A;Cross-references: UNIPROT:Q9HUS4; UNIPARC:UPI00000C5ESF; GB:AE004902; NID:A00000
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4889
C;Superfamily: phthalate dioxygenase reductase

Query Match 51.2%; Score 42; DB 2; Length 366;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PGKKVLGGE 10
|::| |::|
DB 186 PGQRWLGE 194

RESULT 18
E70320
polyribonucleotide nucleotidyltransferase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C;Accession: E70320
R;Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove,

Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70320
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-775 <AOQ>
A;Cross-references: UNIPROT:O66593; UNIPARC:UPI00000562CC; GB:AE000679; NID:g2982936; P1
C;Genetics:
A;Gene: pnpA
C;Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 51.2%; Score 42; DB 2; Length 775;
Best Local Similarity 53.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 GKKVWIGETSSAY 15
| |||||
DB 585 GVRVWVGEGQKVY 597

RESULT 19

CGH2S
collagen alpha 2(I) chain precursor - human
N;Alternate names: procollagen alpha 2(I) chain
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1989 #sequence revision 25-Aug-1995 #text change 31-Dec-2004
C;Accession: A28500; S00824; S09176; I55311; A28472; A42165; A34405; A90567; I55
9005; A02865
R;De Wet, W.; Bernard, M.; Benson-Chanda, V.; Chu, M.L.; Dickson, L.; Weil, D.; Ramirez,
J. Biol. Chem. 262, 16032-16036, 1987
A;Title: Organization of the human pro-alpha-2(I) collagen gene.
A;Reference number: A28500; MUID:88058962; PMID:2824475
A;Accession: A28500
A;Molecule type: DNA; mRNA
A;Residues: 1-248, 'N', 250-1366 <DEM>
A;Cross-references: UNIPROT:P08123; UNIPROT:Q14038; UNIPROT:Q9UMB3; UNIPROT:Q9UMW6; UNIR
R;Kuivaniemi, H.; Tromp, G.; Chu, M.L.; Prockop, D.J.
Biochem. J. 252, 633-640, 1988
A;Title: Structure of a full-length cDNA clone for the prepro-alpha-2(I) chain of human
A;Reference number: S00824; MUID:88339824; PMID:3421913
A;Accession: S00824
A;Molecule type: mRNA
A;Residues: 1-275, 'A', 277-332, 'V', 334-337, 'A', 339-482, 'A', 484-548, 'D', 550-765 <KU11>
A;Cross-references: UNIPARC:UPI000016A6FC; EMBL:Y00724; NID:G30022; PIDN:CAA68709.1; PID
R;Dickson, L.A.; de Wet, W.; di Liberto, M.; Weil, D.; Ramirez, F.
Nucleic Acids Res. 13, 3427-3438, 1985
A;Title: Analysis of the promoter region and the N-propeptide domain of the human proalp
A;Reference number: S09176; MUID:85242047; PMID:4011429
A;Accession: S09176
A;Molecule type: DNA
A;Residues: 1-23, '33-58, 'P', 60-93 <DIC>
A;Cross-references: UNIPARC:UPI0000173B96; UNIPARC:UPI0000173B97; EMBL:X02488; NID:g3009
R;Weil, D.; D'Alessio, M.; Ramirez, F.; Byre, D.R.
J. Biol. Chem. 265, 16007-16011, 1990
A;Title: Structural and functional characterization of a splicing mutation in the pro-al
A;Reference number: I55311; MUID:90368825; PMID:2394758
A;Accession: I55311
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 76-93 <WE11>
A;Cross-references: UNIPARC:UPI000006F17F; GB:M35391; NID:g189684; PIDN:AAA60041.1; PID:
A;Accession: A58111
A;Molecule type: protein
A;Residues: 23-75, 94-96 <WE12>
A;Cross-references: UNIPARC:UPI0000173B98
A;Note: mutant sequence from a patient with Ehlers-Danlos syndrome type VII
R;Wirtz, M.K.; Gianville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, D.W.
J. Biol. Chem. 262, 16376-16385, 1987
A;Title: Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids comprising the N-t
A;Reference number: A28472; MUID:88059013; PMID:3680255
A;Accession: A28472

A;Molecule type: protein
A;Residues: 32-75, 94-111 <WIR>
A;Cross-references: UNIPARC:UPI0000173B99
A;Note: mutant sequence of patient with Ehlers-Danlos syndrome type VIIB
R;Chiodo, A.A.; Hockey, A.; Cole, W.G.
J. Biol. Chem. 267, 6361-6369, 1992
A;Title: A base substitution at the splice acceptor site of intron 5 of the COL1A2 gene i
s-Danlos syndrome type VII.
A;Reference number: A42165; MUID:92210617; PMID:1556139
A;Accession: A42165
A;Molecule type: mRNA
A;Residues: 50-126 <CHI>
A;Cross-references: UNIPARC:UPI0000173B9A
A;Note: parts of this sequence were determined by protein sequencing; a mutant sequence]
R;Weil, D.; D'Alessio, M.; Ramirez, F.; Steinmann, B.; Wirtz, M.K.; Gianville, R.W.; Hol
J. Biol. Chem. 264, 16804-16809, 1989
A;Title: Temperature-dependent expression of a collagen splicing defect in the fibroblast
A;Reference number: A34405; MUID:89380311; PMID:2777808
A;Accession: A34405
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 58-108 <WE13>
A;Cross-references: UNIPARC:UPI0000173B9B; GB:J05049
A;Note: the accession cited by the authors is not found in GenBank
A;Note: parts of this sequence were determined by protein sequencing; a mutant having 93-
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alp
A;Reference number: A90567; MUID:71038625; PMID:5529814
A;Accession: A90567
A;Molecule type: protein
A;Residues: 'Z', '81', 'B', '83-96; 417-447 <CLI>
A;Cross-references: UNIPARC:UPI0000173B9C; UNIPARC:UPI0000173B9D
A;Note: the compositions of peptides CNBR1, CNBR0, and CNBR2 were determined; evidence f
R;Kuivaniemi, H.; Sabol, C.; Tromp, G.; Sippola-Thiele, M.; Prockop, D.J.
J. Biol. Chem. 263, 11407-11413, 1988
A;Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen that ca
is asymptomatic mother.
A;Reference number: I55264; MUID:88298792; PMID:3403536
A;Accession: I55264
A;Status: translation not shown; translated from GB/EMBL/DBJ
A;Molecule type: DNA; mRNA
A;Residues: 145-197 <KU12>
A;Cross-references: UNIPARC:UPI000016AE4F; GB:M21671; NID:g189521; PIDN:AAA59994.1; PID:9
A;Note: single base mutation in intron leads to abnormal splicing of mRNA
R;Chipman, S.D.; Shapiro, J.R.; McKinstry, M.B.; Stover, M.L.; Branson, P.; Rowe, D.W.
J. Bone Miner. Res. 7, 793-805, 1992
A;Title: Expression of mutant alpha (I)-procollagen in osteoblast and fibroblast cultures
A;Reference number: I55485; MUID:92351816; PMID:1642148
A;Accession: I55485
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 163-181, 200-213 <CH2>
A;Cross-references: UNIPARC:UPI0000071E28; GB:S41099; NID:g252702; PIDN:AAB22761.1; PID:9
A;Note: mutant sequence from a patient with osteogenesis imperfecta type IV
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens. I
A;Reference number: A92069; MUID:71001508; PMID:4319110
A;Accession: B92069
A;Molecule type: protein
A;Residues: 175-180 <MOR>
A;Cross-references: UNIPARC:UPI0000014DF12
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Fietzek, P.P.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 47, 257-261, 1974
A;Title: Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-skin
A;Reference number: A91224; MUID:75008198; PMID:4412529
A;Accession: A91224
A;Molecule type: protein
A;Residues: 418-447 <PIE>
A;Cross-references: UNIPARC:UPI00000173B9E

R; Tromp, G.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 5254-5258, 1988
A; Title: Single base mutation in the pro alpha 2(I) collagen gene that causes efficient
A; Reference number: 159125; MUID: 88276936; PMID: 2839839
A; Accession: I59125
A; Status: translation not shown; translated from GB/EMBL/DBDJ
A; Molecule type: DNA
A; Residues: 520-573 <TRO>
A; Cross-references: UNIPARC:UPI0000073009; GB:M21353; NID:g180881; PIDN:AAA52053.1; PID:
A; Note: single base mutation in intron leads to splicing out of exon 28
R; Bernard, M.P.; Myers, J.C.; Chu, M.L.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 1139-1145, 1983
A; Title: Structure of a cDNA for the proalpha-2 chain of human type I procollagen. Compa
A; Reference number: S09174; MUID: 83178919; PMID: 6687691
A; Accession: S09174
A; Molecule type: mRNA
A; Residues: 623-742, 'A', 744-764, 'X', 766-827, 'A', 829-830, 'P', 832-836, 'P', 838-1097, 'L', 109
A; Cross-references: UNIPARC:UPI0000173B9F; GB:J00115; GB:V00503; NID:g30123; PIDN:CAA337
A; Experimental source: skin fibroblast cells
R; Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A; Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A; Reference number: 154365; MUID: 95187161; PMID: 7881420
A; Accession: 168663
A; Status: preliminary; translated from GB/EMBL/DBDJ
A; Molecule type: mRNA
A; Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <FOR>
A; Cross-references: UNIPARC:UPI000006E0A7; GB:L47668; NID:g1009095; PIDN:AAB59577.1; PID
R; Niyibizi, C.; Bonadio, J.; Byers, P.H.; Eyre, D.R.
J. Biol. Chem. 267, 23108-23112, 1992
A; Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) cha
A; Reference number: 155369; MUID: 93054637; PMID: 1385413
A; Accession: 155369
A; Status: translated from GB/EMBL/DBDJ
A; Molecule type: mRNA
A; Residues: 663-666, 'D', 668-670 <NIY>
A; Cross-references: UNIPARC:UPI000011E7D9; GB:L00613; NID:g180888; PIDN:AAB59384.1; PID:
A; Note: mutant sequence from a patient with osteogenesis imperfecta
R; Bateman, J.F.; Hannagan, M.; Chan, D.; Cole, W.G.
Biochem. J. 276, 765-770, 1991
A; Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitut
e method.
A; Reference number: A56799; MUID: 91291136; PMID: 2064612
A; Accession: A56799
A; Molecule type: mRNA
A; Residues: 672-675, 'V', 677, 'P', 679-681 <BAT>
A; Cross-references: UNIPARC:UPI000016B384; GB:S39878; NID:g1679911; PIDN:AAB19314.1; PID:
A; Note: sequence extracted from NCBI backbone (NCBIN:39878, NCBI:P:39886)
A; Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors sug
ntrol sequence
R; Maekelae, J.K.; Vuorio, T.; Vuorio, E.
Biochim. Biophys. Acta 1049, 171-176, 1990
A; Title: Growth-dependent modulation of type I collagen production and mRNA levels in cu
A; Reference number: S10768; MUID: 90304220; GB:J00114; NID:g180393; PIDN:AAA51996.1; PID:
A; Accession: S10768
A; Molecule type: mRNA
A; Residues: 960-1021, 'L', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAE>
A; Cross-references: UNIPARC:UPI000016A7ID; EMBL:X55525; NID:g30101; PIDN:CAA39142.1; PID:
A; Experimental source: fibroblast cell culture
R; Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.
Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981
A; Title: Cloning a cDNA for the pro-alpha2 chain of human type I collagen.
A; Reference number: A18855; MUID: 81273090; PMID: 6267597
A; Accession: A18855
A; Molecule type: mRNA
A; Residues: 964-979, 'V', 981-1018, 'Q', 1020 <MYE>
A; Cross-references: UNIPARC:UPI0000173BA0; GB:J00114; NID:g180393; PIDN:AAA51996.1; PID:
A; Note: 1019-Leu was also found
R; Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.
J. Biol. Chem. 263, 7734-7740, 1988
A; Title: Arginine for glycine substitution in the triple-helical domain of the products
A; Reference number: 155285; MUID: 88227975; PMID: 2897363

A:Accession: 155285
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1090-1107 <WEN1>
A:CROSS-references: UNIPARC:UPI000016A612; NID:g179602; PIDN:AAAS1844.1; PID:
A:Accession: I70059
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1090-1101,'R',1103-1107 <WEN2>
A:CROSS-references: UNIPARC:UPI000016A613; GB:M22817; NID:g179606; PIDN:AAAS1846.1; PID:
A>Note: mutant sequence from a patient with osteogenesis imperfecta type IV

Query Match 51.2%; Score 42; DB 1; Length 1366;
Best Local Similarity 72.7%; Pred. No. 1.le+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 KKVWLGETSSA 14
Db 1224 KHVWLGETINA 1234
| | | | | : |
| | | | | : |

RESULT 20
T02433
DNA binding protein EREBP-3 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02433
R:Ohme-Takagi, M.; Shinshi, H.
Plant Cell 7, 173-182, 1995
A>Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsive
A:Reference number: Z14671; MUID:95276459; PMID:7756828
A:Accession: T02433
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-225 <OHM>
A:CROSS-references: UNIPROT:Q04077; UNIPARC:UPI000009FB50; EMBL:D38124; NID:g790360; PID:
A:Experimental source: strain Bv4; tissue-type leaf

Query Match 50.0%; Score 41; DB 2; Length 225;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 11; Conservative 1; Mismatches 2; Indels 8; Gaps 2;

Qy 2 PGKK--VWLG-----ETSSAY 15
| | | | | : |
| | | | | : |
Db 45 PGKSRVLGLGTFTAEEAKAY 65

RESULT 21
AH2866
S-methyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AH2866
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2866
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <KUR>
A:CROSS-references: UNIPROT:Q8UCX1; UNIPARC:UPI00000D1E8F; GB:AE008688; PIDN:AAL43350.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2362
A:Map position: circular chromosome

Query Match 50.0%; Score 41; DB 2; Length 306;
Best Local Similarity 66.7%; Pred. No. 39;

```

Matches      8;  Conservative      0;  Mismatches      4;  Indels      0;  Gaps      0;

Qy      2  PGKKVWLGETSS 13
      |  |||||  |
Db      135  PSVDVWLGETLS 146

RESULT 22
E97643
msh protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: E97643
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E97643
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <KUR>
A:Cross-references: UNIPROT:Q8UCX1; UNIPARC:UPI000000D1B8F; GB:AE007869; PIDN:AAK88102.1
C:Genetics:
A:Gene: AGR_C 4290
A:Map position: circular chromosome

      Query Match      50.0%;  Score 41;  DB 2;  Length 306;
      Best Local Similarity 66.7%;  Pred. No. 39;
      Matches      8;  Conservative      0;  Mismatches      4;  Indels      0;  Gaps      0;

Qy      2  PGKKVWLGETSS 13
      |  |||||  |
Db      135  PSVDVWLGETLS 146

RESULT 23
S36444
hygromycin phosphotransferase - Pseudomonas pseudomallei
C:Species: Pseudomonas pseudomallei
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:Accession: S36444
R:Penalzoza-Vazquez, A.; Herrera-Estrella, L.; Bailey, A.M.
submitted to the EMBL Data Library, July 1993
A:Description: Cloning and sequencing of the genes involved in glyphosate catabolism by
A:Reference number: S36444
A:Accession: S36444
A:Molecule type: DNA
A:Residues: 1-420 <PEN>
A:Cross-references: UNIPROT:Q52501; UNIPARC:UPI000000B1D35; EMBL:X74325; NID:g439726; PID
A:Experimental source: strain 22

      Query Match      50.0%;  Score 41;  DB 2;  Length 420;
      Best Local Similarity 53.8%;  Pred. No. 53;
      Matches      7;  Conservative      1;  Mismatches      5;  Indels      0;  Gaps      0;

Qy      2  PGKKVWLGETSS 14
      ||  ||  ||  ||  ||
Db      383  PGPKTWAGDSQA 395

RESULT 24
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pa
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family m
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA

```

A;Residues: 1-480 <MCK>
A;Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E89A; GB:AF282885
C;Comment: This protein, a intracellular membrane-bound enzyme, has biological and therapeutic.
C;Genetics:
A;Gene: hpa2a
A;Map position: 10q23-10q24
C;Keywords: heparin binding; membrane bound

Query Match 50.0%; Score 41; DB 2; Length 480;
Best Local Similarity 85.7%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PGKKVWL 8
|||:|
Db 261 PGKKIWL 267

RESULT 25
AH2248
proteinase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH2248
R;kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AH2248
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-615 <KUR>
A;Cross-references: UNIPROT:Q8VRAS; UNIPARC:UPI00000CE81A; GB:BA000019; PIDN:BA075242.1;
C;Genetics:
A;Gene: alr3543

Query Match 50.0%; Score 41; DB 2; Length 615;
Best Local Similarity 77.8%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PGKKVWLGE 10
|||:|
Db 112 PAKPVWLGE 120

RESULT 26
AD0125
exodeoxyribonuclease V (EC 3.1.11.5) beta chain [imported] - Yersinia pestis (strain C09)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AD0125
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, F.; Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AD0125
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1220 <KUR>
A;Cross-references: UNIPROT:Q8ZH88; UNIPARC:UPI00000CD786; GB:AL590842; PIDN:CAC89863.1;
C;Genetics:
A;Gene: recB
C;Superfamily: exodeoxyribonuclease V 135K chain

Query Match 50.0%; Score 41; DB 2; Length 1220;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;


```

QY 4 KKWVLGETSSAY 15
| | | | |
Db 1103 KSNWLGEDSRAY 1114

RESULT 27
AE3301
C:Species: Brucella melitensis (strain 122)
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AE3301
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
.; Mazur, M.; Golezman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesee
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AE3301
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KUR>
A:Cross-references: UNIPROT:Q8YIP8; UNIPARC:UPI0000057C76; GB:AE008917; PIDN:AAL51576.1;
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BM3I0395
A:Map position: 1
C:Keywords: oxidoreductase

Query Match 48.8%; Score 40; DB 2; Length 122;
Best Local Similarity 58.3%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14
| | | | |
Db 66 GKKAWAGEKGEA 77

RESULT 28
S64937
probable membrane protein YLR101c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L2705
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C:Accession: S64937
R:Messenguy, F.; Dubois, E.; Vierendeels, F.; Scherens, B.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64935
A:Accession: S64937
A:Molecule type: DNA
A:Residues: 1-131 <MES>
A:Cross-references: UNIPROT:Q08027; UNIPARC:UPI000006C33C; EMBL:Z73272; NID:gi360482; PI
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YLR101c
A:Cross-references: SGD:S0004091
A:Map position: 12R
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YLR101c
C:Keywords: transmembrane protein
F:45-62/Domain: transmembrane #status predicted <TMM>

Query Match 48.8%; Score 40; DB 2; Length 131;
Best Local Similarity 54.5%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KWLGETSSAY 15
| | | | |
Db 96 KMWIGQSSYYV 106

RESULT 29
TS2019
ethylene responsive element binding factor 4 [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004

```

```

C:Accession: TS2019
R:Fujimoto, S.Y.; Ohta, M.; Usui, A.; Shinshi, H.; Ohme-Takagi, M.
Plant Cell 12, 393-404, 2000
A:Title: Arabidopsis ethylene responsive element binding factors act as transcriptional
A:Reference number: Z25893
A:Accession: TS2019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-222 <FUG>
A:Cross-references: UNIPROT:O80340; UNIPARC:UPI000000140F; EMBL:AB008106; PIDN:BAA32421.1
C:Genetics:
A:Gene: ERF-4
C:Function:
A:Description: can act as transcriptional repressor in leaves [validated, MUID:20181733]

Query Match 48.8%; Score 40; DB 2; Length 222;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 11; Conservative 1; Mismatches 2; Indels 8; Gaps 2;

QY 2 PGKK--VWLG-----ETSSAY 15
| | | | |
Db 43 PGKTRVWLGTFTAEAAAY 64

RESULT 30
G97553
sugar fermentation stimulation protein homolog [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: G97553
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G97553
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <KUR>
A:Cross-references: UNIPROT:P58429; UNIPARC:UPI00001358C6; GB:AE007869; PIDN:AAK87384.1;
C:Genetics:
A:Gene: AGR_C 2956
A:Map position: circular chromosome
C:Superfamily: sugar fermentation stimulation protein

Query Match 48.8%; Score 40; DB 2; Length 239;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13
| | | | |
Db 47 PGSRWLSEHDS 58

RESULT 31
AI2773
sugar fermentation stimulation protein [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Kap, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AI2773
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <KUR>
A:Cross-references: UNIPROT:P58429; UNIPARC:UPI00001358C6; GB:AE008688; PIDN:AAL42607.1;

```

```

A;Experimental source: strain C58 (Dupont)
C;Gene: sfsa
A;Map position: circular chromosome
C;Superfamily: sugar fermentation stimulation protein

Query Match      48.8%; Score 40; DB 2; Length 239;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13
    ||:|||||
Db 47 PGSRIWLSEHDS 58

RESULT 32
B82979
hypothetical protein PA5340 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: B82979
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B82979
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <STO>
A;Cross-references: UNIPROT:Q9HTL6; UNIPARC:UPI00000C5FB0; GB:AE004946; GB:AE004091; NID:10382966
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA5340

Query Match      48.8%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PGKKVWLGE 10
    ||:|||||
Db 234 PGKRVWLLE 242

RESULT 33
E72623
probable autoantigen APE1445 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 05-Oct-2004
C;Accession: E72623
R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kikuchi, K.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: E72623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <RAW>
A;Cross-references: UNIPROT:Q9YC05; UNIPARC:UPI000005DPF8; DDBJ:AP0000061; NID:gs104821;
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1445
C;Superfamily: Exosome complex 3'-5' exonuclease

Query Match      48.8%; Score 40; DB 2; Length 276;
Best Local Similarity 75.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKVWL 8
    ||:|||||
Db 132 RPGEKVV 139

```

RESULT 34

T02434

DNA binding protein EREBP-4 - common tobacco

C;Species: Nicotiana tabacum (common tobacco)

C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 31-Dec-2004

C;Accession: T02434

R;Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsive element

A;Reference number: Z14671; MUID:95276459; PMID:7756828

A;Accession: T02434

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-291 <OHM>

A;Cross-references: UNIPROT:Q40478; UNIPARC:UPI000009DFAD; EMBL:D38125; NID:g790361; PID:

A;Experimental source: strain BY4; tissue-type leaf

Query Match 48.8%; Score 40; DB 2; Length 291;

Best Local Similarity 42.9%; Pred. No. 54;

Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

QY 1 RPKKVWLGE-----ETSSAY 15

Db 165 RKGTRVWLGTPTTAIEAAKAY 185

RESULT 35

T09544

phosphoprotein phosphatase (EC 3.1.3.16), catalytic beta chain - alfalfa

C;Species: Medicago sativa (alfalfa)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C;Accession: T09544

R;Visi, E.; Csordas Toth, E.; Kovacs, I.; Magyar, G.; Horvath, G.; Bagossi, P.; Gergely,

Arch. Biochem. Biophys. 360, 206-214, 1998

A;Title: Protein phosphatase 1 catalytic subunit isoforms from alfalfa: Biochemical characterization

A;Reference number: Z16730; MUID:99068922; PMID:9851832

A;Accession: T09544

A;Molecule type: mRNA

A;Residues: 1-326 <VIS>

A;Cross-references: UNIPROT:O65844; UNIPARC:UPI00000A4B32; EMBL:AJ002485; NID:g3176071; I:

C;Genetics:

A;Gene: PPI beta

C;Function:

A;Description: catalyzes hydrolysis of peptidyl-phosphoserine or -phosphothreonine to release

C;Superfamily: serine/threonine protein phosphatase; phosphoesterase core homology; phospho

C;Keywords: phosphoric monoester hydrolase; serine/threonine-specific phosphatase

F;26-285/Domain: phosphoprotein phosphatase homology <PPP>

F;54-122/Domain: phosphoesterase core homology <PEC>

Query Match 48.8%; Score 40; DB 2; Length 326;

Best Local Similarity 72.7%; Pred. No. 60;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPKKVWLGET 11

Db 19 RPGEKVLSET 29

RESULT 36

JC5191

contractile tail sheath protein - Pseudomonas aeruginosa phage PS17

C;Species: Pseudomonas aeruginosa phage PS17

C;Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004

C;Accession: JC5191; JC4865

R;Sasaki, T.; Shinomiya, T.; Kumazaki, T.; Mohri, N.; Ishii, S.; Arisaka, F.

Res. Commun. Biochem. Cell Mol. Biol. 1, 93-107, 1997

A;Title: Nucleotide sequences of the contractile tail sheath and tube genes of bacteriophage

A;Reference number: JC5191

A;Accession: JC5191

A;Molecule type: DNA

A;Residues: 1-386 <SA2>

A;Cross-references: UNIPROT:Q38068; UNIPARC:UPI00009BCD6; DDBJ:D26449; NID:g452162; PID
C;Genetics:
A;Gene: FI
C;Keywords: tail protein
F:2-386/Product: tail sheath protein #status predicted <MAT>

Query Match 48.8%; Score 40; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 71;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PGKKVWLGETSS 13
||:|||||
Db 183 PGQVWNTETSS 194

RESULT 37
S73012
polyketide synthase pksB - Mycobacterium leprae
N;Alternate names: L518 Fl 1 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S73012
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid L518.
A;Reference number: S72591
A;Accession: S73012
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-958 <SMI>
A;Cross-references: UNIPROT:Q49931; UNIPARC:UPI00000BD072; EMBL:U00023; NID:g467194; PID
C;Genetics:
A;Start codon: GTG
C;Keywords: carrier protein
F:572-767/Domain: short-chain alcohol dehydrogenase homology <SAD2>
F:843-914/Domain: acyl carrier protein homology <ACP1>

Query Match 48.8%; Score 40; DB 2; Length 958;
Best Local Similarity 60.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KKWLGTTSS 13
||:|||||
Db 286 KQVWLGTTAT 295

RESULT 38
F96501
hypothetical protein F28H19.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: F96501
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F96501
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1014 <STO>
A;Cross-references: UNIPROT:Q9MAR9; UNIPARC:UPI00000A7CAF; GB:AE005173; NID:g7523671; PI
C;Genetics:
A;Gene: F28H19.4
A;Map position: 1

Query Match 48.8%; Score 40; DB 2; Length 1014;

Best Local Similarity 41.7%; Pred. No. 1.8e+02;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PGKKVWLGETSS 13
||:|||||
Db 843 PGQIWMGKSDS 854

RESULT 39
WMFM12
125K protein - alfalfa mosaic virus
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (EC
C;Species: alfalfa mosaic virus, AMV
C;Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 09-Jul-2004
C;Accession: A04197
R;Cornelissen, B.J.C.; Brederode, F.T.; Moormann, R.J.M.; Bol, J.F.
Nucleic Acids Res. 11, 1253-1265, 1983
A;Title: Complete nucleotide sequence of alfalfa mosaic virus RNA 1.
A;Reference number: A04197; MUID:83143345; PMID:6298738
A;Accession: A04197
A;Molecule type: mRNA
A;Residues: 1-1126 <COR>
A;Cross-references: UNIPROT:P03589; UNIPARC:UPI0000137F7B; GB:L00163; GB:J02000; NID:g333
C;Genetics:
A;Map position: segment 1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; mRNA capping; nucleotide binding; P-loop; S-ade
F:99-188/Domain: methyltransferase #status predicted <MTF>
F:838-845/Region: nucleotide-binding motif A (P-loop)
F:845-1036/Domain: helicase #status predicted <HHS>

Query Match 48.8%; Score 40; DB 1; Length 1126;
Best Local Similarity 42.9%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PGKKVWLGETSSAY 15
||:|||||
Db 752 PKXNTWVGPTARSY 765

RESULT 40
A87204
polyketide synthase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: A87204
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: A87204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: UNIPROT:Q9Z5K6; UNIPARC:UPI00000D42DC; GB:AL450380; NID:g13093966; PJ
C;Genetics:
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoacy
homology; [acyl-carrier-protein] S-malonyltransferase homology
C;Keywords: carrier protein

Query Match 48.8%; Score 40; DB 2; Length 1871;
Best Local Similarity 60.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KKWLGTTSS 13
||:|||||
Db 1199 KQVWLGTTAT 1208

```
RESULT 41
AB2718
conserved hypothetical protein Atu1147 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2718
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, S.; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2718
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-115 <KUR>
A:Cross-references: UNIPROT:Q8UG92; UNIPARC:UPI000016459D; GB:AE008688; PIDN:AAL42160.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu1147
A:Map position: circular chromosome
C:Superfamily: Bacillus subtilis Hypothetical protein yung

Query Match 47.6%; Score 39; DB 2; Length 115;
Best Local Similarity 58.3%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVLGETSSAY 15
||| ||||| :
DB 13 KKGWSGETSNLW 24

RESULT 42
F97499
hypothetical protein AGR_C_2123 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 31-Dec-2004
C:Accession: F97499
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F97499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: UNIPROT:Q8UG92; UNIPARC:UPI00000DIA64; GB:AE007869; PIDN:AAK86951.1;
C:Genetics:
A:Gene: AGR_C_2123
A:Map position: circular chromosome
C:Superfamily: Bacillus subtilis hypothetical protein yung

Query Match 47.6%; Score 39; DB 2; Length 182;
Best Local Similarity 58.3%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVLGETSSAY 15
||| ||||| :
DB 80 KKGWSGETSNLW 91

RESULT 43
A86461
hypothetical protein F14W2.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A86461
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
```

```
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzbeg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86461
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-184 <STO>
A:Cross-references: UNIPROT:Q9LQ28; UNIPARC:UPI00000A7C94; GB:AE005172; NID:g9665094; PII
C:Genetics:
A:Map position: 1

Query Match 47.6%; Score 39; DB 2; Length 184;
Best Local Similarity 53.3%; Pred. No. 51;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 2 PGKK--VWLGETSSA 14
||| :||| :
DB 38 PGKTRIWLSVETA 52

RESULT 44
H64968
acetyl CoA acetyltransferase - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: H64968; I69646; I69656
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64968
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <BLAT>
A:Cross-references: UNIPROT:P37750; UNIPARC:UPI000003EB11; GB:AE000294; GB:U00096; NID:G
R:Yao, Z.; Valvano, M.A.
J. Bacteriol. 176, 4133-4143, 1994
A:Title: Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb)
erotypes Y and 4a.
A:Reference number: I55053; MUID:94292434; PMID:7517390
A:Accession: I69646
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-167, 'LPRKYCHC', 177-186, 'IMR', 190-196 <RES>
A:Cross-references: UNIPARC:UPI000016F12B; EMBL:U03041; NID:g501028; PIDN:AAC31635.1; PII
R:Stevenson, G.; Neal, B.; Liu, D.; Hobbs, M.; Packer, N.H.; Batley, M.; Redmond, J.W.; I
J. Bacteriol. 176, 4144-4156, 1994
A:Title: Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb
A:Reference number: I55054; MUID:94292435; PMID:7517391
A:Accession: I69656
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-167, 'LPRKYCHC', 177-186, 'IMR', 190-196 <RE2>
A:Cross-references: UNIPARC:UPI000016F12B; EMBL:U09876; NID:g508236; PID:g508245
C:Genetics:
A:Gene: yefH
A:Map position: 45 min
C:Superfamily: galactoside acetyltransferase

Query Match 47.6%; Score 39; DB 2; Length 196;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 GKQVWLGE 10
||| :|||
DB 136 GQVWLGE 143
```

```
RESULT 45
T17997
hypothetical protein A495R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17997
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17997
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-221 <GRA>
A:Cross-references: UNIPROT:Q98545; UNIPARC:UPI00000F8EBB; EMBL:U42580; NID:g4028896; P1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A495R
C:Superfamily: Chlorella virus PBCV-1 hypothetical protein A315L
Query Match 47.6%; Score 39; DB 2; Length 221;
Best Local Similarity 53.8%; Pred. No. 60;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 108 KKGNKNWLGKTHS 120
:|:|:|:|:|:|
1 RPKKKVWLGETSS 13
:|:|:|:|:|:|
108 KKGNKNWLGKTHS 120

RESULT 46
F64901
ABC-type transport protein b1483 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 05-Oct-2004
C:Accession: F64901
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64901
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <BLAT>
A:Cross-references: UNIPROT:P77622; UNIPARC:UPI000013A940; GB:AE000245; GB:U000096; NID:g
A:Experimental source: strain K-12, substrain MGL655
C:Keywords: ATP; nucleotide binding; P-loop
F;32-219/Domain: ATP-binding cassette homology <ABC>
F;49-56/Region: nucleotide-binding motif A (P-loop)
Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 2 PGKKVWLGETS 12
:|:|:|:|:|:|
16 PARKNWLKTT 26

RESULT 47
B85728
hypothetical protein Z2227 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 05-Oct-2004
C:Accession: B85728
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85728
A>Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-308 <STO>
A:Cross-references: UNIPROT:Q8X4Z6; UNIPARC:UPI00000D0D46; GB:AE005174; NID:g12515198; P1
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2227
Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 2 PGKKVWLGETS 12
:|:|:|:|:|:|
16 PARKNWLKTT 26

RESULT 48
G90889
hypothetical protein ECs2087 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 05-Oct-2004
C:Accession: G90889
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90889
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <HAY>
A:Cross-references: UNIPROT:Q8X4Z6; UNIPARC:UPI00000D0D46; GB:BA000007; PIDN:BAB35510.1;
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2087
Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 2 PGKKVWLGETS 12
:|:|:|:|:|:|
16 PARKNWLKTT 26

RESULT 49
XNBYUG
UDPglucose-hexose-1-phosphate uridylyltransferase (EC 2.7.7.12) - yeast (Saccharomyces ce
N:Alternate names: galactose-1-phosphate uridylyltransferase; protein YBR018c; protein YBR
C:Species: Saccharomyces cerevisiae
C>Date: 17-Mar-1987 #sequence_revision 09-Sep-1994 #text_change 05-Oct-2004
C:Accession: S45873; S50813; S05811; A00720; S18725; S18757; S24918; S50322
R:Entian, K.D.; Koetter, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; Bolet
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45862
A:Accession: S45873
A:Molecule type: DNA
A:Residues: 1-366 <ENT>
A:Cross-references: UNIPROT:P08431; UNIPARC:UPI00001682B8; EMBL:Z35887; NID:g536219; PID
A:Experimental source: strain S288C
R:Schaeff-Gerstenschlaeger, I.; Schindewolf, T.; Lehnert, W.; Rose, M.; Zimmermann, F.K.
Yeast 11, 79-83, 1995
A>Title: Sequence and functional analysis of a 7.2 kb fragment of Saccharomyces cerevisiae
A:Reference number: S50812; MUID:95282516; PMID:7762304
A:Accession: S50813
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-366 <SCW>
A:Cross-references: UNIPARC:UPI00001682B8; EMBL:X81324; NID:g587572; PIDN:CAA57105.1; P1
A:Experimental source: strain S288C
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994
R:Tajima, M.; Nogi, Y.; Fukaeawa, T.
Yeast 1, 67-77, 1985
```

A;Title: Primary structure of the *Saccharomyces cerevisiae* GAL7 gene.
 A;Reference number: S05811; MUID:89131252; PMID:2851900
 A;Accession: S05811
 A;Molecule type: DNA
 A;Residues: 1-84; 'S', 87-266, 'A', 268-344, 'I', 346-366 <TAJ>
 A;Cross-references: UNIPARC:UPI0000168C2E; EMBL:M12348; NID:gl71559; PIDN:AAA34627.1; PI
 R;Citron, B.A.; Donelson, J.E.
 J. Bacteriol. 158, 269-278, 1984
 A;Title: Sequence of the *Saccharomyces* GAL region and its transcription in vivo.
 A;Reference number: A91795; MUID:84185433; PMID:6715281
 A;Accession: A00720
 A;Molecule type: DNA
 A;Residues: 1-10, 'Y', 12-57, 'H', 59-184 <CIT>
 A;Cross-references: UNIPARC:UPI0000172706; EMBL:X01752; NID:gl71561; PIDN:AAA34628.1; PI
 R;Nogi, Y.; Fukasawa, T.
 Nucleic Acids Res. 11, 8555-8568, 1983
 A;Title: Nucleotide sequence of the transcriptional initiation region of the yeast GAL7
 A;Reference number: S18725; MUID:84169499; PMID:6324089
 A;Accession: S18725
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-21 <NOG>
 A;Cross-references: UNIPARC:UPI0000172707; EMBL:X00215
 A;Accession: S18757
 A;Molecule type: protein
 A;Residues: 2-8 <NOG2>
 A;Cross-references: UNIPARC:UPI0000172708
 C;Genetics:
 A;Gene: SGD:GAL7; MIPS:YBR018C
 A;Cross-references: SGD:S0000222; MIPS:YBR018C
 A;Map position: 2R
 C;Function:
 A;Description: galactose metabolism; nucleotidyltransferase
 C;Superfamily: galactose-1-phosphate uridylyltransferase
 C;Keywords: galactose metabolism; nucleotidyltransferase
 F;2-366/Product: UDPglucose-hexose-1-phosphate uridylyltransferase #status experimental

Query Match 47.6%; Score 39; DB 1; Length 366;
 Best Local Similarity 50.0%; Pred. No. 99;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKKVWLGETSSAY 15
 | : ||| : ||
 DB 32 KRPWLGGQEAAY 43

RESULT 50
 B70579
 probable cell division protein FtsZ - *Mycobacterium tuberculosis* (strain H37RV)
 C;Species: *Mycobacterium tuberculosis*
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C;Accession: B70579
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Sgares, R.; Sulstson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9634230
 A;Accession: B70579
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-379 <COL>
 A;Cross-references: UNIPROT:O08378; UNIPARC:UPI000012AD2A; GB:Z95388; GB:AL123456; NID:9
 C;Genetics:
 A;Gene: ftsZ
 C;Superfamily: cell division protein ftsZ

Query Match 47.6%; Score 39; DB 2; Length 379;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
 ||:|:|:|:|:|:
 DB 317 PGRKPVWGETGGAH 330

Search completed: June 5, 2006, 12:53:10
 Job time : 44.6438 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 131.507 Seconds

(without alignment)
105.510 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPCKKWLGETSSAY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : UniProt 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	543	1 HPSE_HUMAN	Q9Y251 homo sapien
2	82	100.0	558	2 Q333X5_SPAJD	Q333x5 spalax juda
3	82	100.0	574	2 Q333X6_SPAJD	Q333x6 spalax juda
4	82	100.0	574	2 Q333X7_SPAJD	Q333x7 spalax juda
5	82	100.0	574	2 Q333X8_SPAJD	Q333x8 spalax juda
6	82	100.0	574	2 Q333X9_SPAJD	Q333x9 spalax juda
7	77	93.9	535	1 HPSE_MOUSE	Q6Ygz1 mus musculus
8	77	93.9	535	1 HPSE_CHICK	Q71rpl rattus norv
9	73	89.0	523	1 HPSE_RAT	Q9ybk5 gallus gall
10	71	86.6	533	2 Q4SYF6_TETNG	Q4syf6 tetraodon n
11	68	82.9	545	1 HPSE_BOVIN	Q9myy0 bos taurus
12	64	78.0	255	2 Q4TGC8_TETNG	Q4tgc8 tetraodon n
13	50	61.0	597	2 Q4TB80_TETNG	Q4tb80 tetraodon n
14	49	59.8	586	2 Q21PR2_9DEL	Q21pr2 anaeromyxob
15	48	58.5	197	2 Q36KN3_WARHY	Q36kn3 marinobacte
16	48	58.5	616	2 Q3QDK4_9GAMM	Q3qdk4 shewanella
17	47	57.3	334	2 Q41J33_METBU	Q41j33 methanococc
18	47	57.3	426	2 Q54359_STRLN	Q54359 streptomyce
19	47	57.3	3143	2 Q6LAL3_9POTV	Q6lal3 plum pox vi
20	47	57.3	3143	2 Q6V3X4_9POTV	Q6v3x4 plum pox vi
21	46	56.1	64	2 Q858F7_9CAUD	Q858f7 enterobacte
22	46	56.1	157	2 Q6J9N4_ARATH	Q6j9n4 arabidopsi
23	46	56.1	203	2 Q47T58_THEFY	Q47t58 thermobifid
24	46	56.1	827	2 Q7S5G5_NEUCR	Q7s5g5 neurospora
25	46	56.1	1180	1 EX5B_ECOLI	Q08394 escherichia
26	46	56.1	1180	2 Q31XG8_SHIDS	Q31xg8 shigella bo
27	46	56.1	1180	2 Q32CAI_SHIDS	Q32cai shigella dy
28	46	56.1	1180	2 Q3Y440_SHIDS	Q3y440 shigella so
29	46	56.1	1180	2 Q2MAL7_ECOLI	Q2mal7 escherichia
30	46	56.1	1180	2 Q8X6M9_ECO57	Q8x6m9 escherichia
31	46	56.1	1180	2 Q83JW0_SHIFL	Q83jw0 shigella fl

32	46	56.1	1183	2 Q8FEB3_ECOL6	Q8feb3 escherichia
33	45.5	55.5	1596	2 Q8YKV0_ANASP	Q8ykv0 anabaena sp
34	45	54.9	73	2 Q3PI29_PARDE	Q3pi29 paracoccus
35	45	54.9	134	2 Q41177_SYNP2	Q41177 synchococc
36	45	54.9	164	1 DRE2H_ARATH	DRE2H arabidopsi
37	45	54.9	225	2 Q5BK39_RAT	Q5bk39 rattus norv
38	45	54.9	225	2 Q6P3A6_MOUSE	Q6p3a6 mus musculu
39	45	54.9	225	2 Q8C708_MOUSE	Q8c708 m o day neo
40	45	54.9	226	2 Q8A4U3_BACTN	Q8a4u3 bacteroides
41	45	54.9	331	2 Q3ITU3_NATPD	Q3itu3 natronomona
42	45	54.9	335	2 Q6FA89_ACTAD	Q6fa89 acinetobact
43	45	54.9	473	2 Q5XK86_XENLA	Q5xk86 xenopus lae
44	45	54.9	562	2 Q57MD5_SALCH	Q57md5 salmonella
45	45	54.9	562	2 Q5PE41_SALPA	Q5pe41 salmonella
46	45	54.9	562	2 Q8ZNK5_SALTY	Q8znk5 salmonella
47	45	54.9	562	2 Q8Z592_SALTI	Q8z592 salmonella
48	45	54.9	962	2 Q5SE54_DICDI	Q5se54 dictyosteli
49	45	54.9	1869	2 Q3HF42_TRIER	Q3hf42 trichodesmi
50	44	53.7	197	2 Q36TJ3_WARHY	Q36tj3 marinobacte
51	44	53.7	246	2 Q4UIV3_9BACL	Q4ulv3 paenibacill
52	44	53.7	329	2 Q6LKM9_PROPR	Q6lkm9 photobacter
53	44	53.7	336	2 Q9SKT1_ARATH	Q9skt1 arabidopsi
54	44	53.7	341	2 Q6MZF8_HUMAN	Q6mzf8 homo sapien
55	44	53.7	349	2 Q5ARU6_EMENI	Q5aru6 aspergillus
56	44	53.7	447	2 Q5FUI5_GLUOX	Q5fui5 gluconobact
57	44	53.7	482	1 MTTB_METTE	Q20995 methanobarc
58	44	53.7	563	1 PTFBC_ECOLI	P20966 e pts syste
59	44	53.7	563	2 Q31YX5_SHIDS	Q31yx5 shigella bo
60	44	53.7	563	2 Q32E63_SHIDS	Q32e63 shigella dy
61	44	53.7	563	2 Q32039_SHISS	Q32039 shigella so
62	44	53.7	563	2 Q8XE91_ECO57	Q8xe91 escherichia
63	44	53.7	563	2 Q8FFT4_ECOL6	Q8fft4 escherichia
64	44	53.7	563	2 Q83QM2_SHIFL	Q83qm2 shigella fl
65	44	53.7	592	2 Q6P3X8_HUMAN	Q6p3x8 homo sapien
66	44	53.7	810	2 Q2UC74_ASPOR	Q2uc74 aspergillus
67	44	53.7	819	2 Q418Y0_GIBZE	Q418y0 gibberella
68	44	53.7	834	2 Q55J37_CRYNE	Q55j37 cryptococcu
69	44	53.7	834	2 Q5KA04_CRYNE	Q5ka04 cryptococcu
70	44	53.7	893	2 Q9Y1Y3_9METZ	Q9y1y3 ephydatia f
71	44	53.7	1016	2 Q5FT91_GLUOX	Q5ft91 gluconobact
72	43.5	53.0	382	2 Q4K3M2_PSEF5	Q4k3m2 pseudomonas
73	43.5	53.0	1187	2 Q93284_FUGRU	Q93284 fugu rubrip
74	43	52.4	249	2 Q9KGF2_BACHD	Q9kgf2 bacillus ha
75	43	52.4	255	2 Q445E2_SOLUS	Q445e2 solibacter
76	43	52.4	304	2 Q7QTN7_GIALA	Q7qtn7 giardia lam
77	43	52.4	331	2 Q4IGM6_GIBZE	Q4igm6 gibberella
78	43	52.4	385	2 Q47QC3_THEFY	Q47qc3 thermobifid
79	43	52.4	386	2 Q5DBH7_SCHJA	Q5dbh7 schistosoma
80	43	52.4	410	2 Q7NZY8_CHRVO	Q7nzy8 chromobacte
81	43	52.4	469	2 Q8UWV9_9INFA	Q8uwv9 influenza a
82	43	52.4	469	2 Q2IDF0_9INFA	Q2idf0 influenza a
83	43	52.4	494	1 MTTB1_METAC	Q8tta9 methanobarc
84	43	52.4	494	1 MTTB1_METBA	Q93558 methanobarc
85	43	52.4	494	1 MTTB_METBF	FUC0W7 methanobarc
86	43	52.4	494	1 Q95AF6_9ASPA	Q95af6 caesia cont
87	43	52.4	495	1 MTTB1_METNA	P58973 methanobarc
88	43	52.4	495	1 MTTB2_METAC	Q8ts73 methanobarc
89	43	52.4	495	1 MTTB2_METNA	P58974 methanobarc
90	43	52.4	766	2 Q4DKD9_TRYCR	Q4dkd9 trypanosoma
91	43	52.4	766	2 Q4CVY5_TRYCR	Q4cvy5 trypanosoma
92	43	52.4	871	2 Q37YK6_SPHAR	Q37yk6 novosphingo
93	43	52.4	877	2 Q2UGU9_ASPOR	Q2ugu9 aspergillus
94	43	52.4	987	2 Q6NRB9_XENLA	Q6nrb9 xenopus lae
95	43	52.4	1022	2 Q9LD58_ARATH	Q9ld58 arabidopsi
96	43	52.4	1022	2 Q9LET7_ARATH	Q9let7 arabidopsi
97	43	52.4	1180	2 Q2NRH7_SODGL	Q2nrh7 sodalis glo
98	43	52.4	1181	2 Q57KC4_SALCH	Q57kc4 salmonella
99	43	52.4	1181	2 Q5PEM8_SALPA	Q5pem8 salmonella
100	43	52.4	1181	2 Q8ZMB6_SALTY	Q8zmb6 salmonella
101	43	52.4	1181	2 Q8Z419_SALTI	Q8z419 salmonella
102	43	52.4	1536	2 Q8GAL7_DICDI	Q8gal7 dictyosteli
103	43	52.4	2767	2 Q4B4A1_9BURK	Q4b4a1 polaromonas
104	42.5	51.8	580	2 Q3M2M4_ANAVT	Q3m2m4 anabaena va

105	42	51.2	73	2	Q44NM0_CHLLI	Q44nm0 chlorobium	178	41	50.0	365	2	Q6K7E6_ORYSA	Q6k7e6 oryza sativ
106	42	51.2	113	2	Q6I823_OSCAG	Q6i823 oscillatori	179	41	50.0	365	2	Q9SE28_ORYSA	Q9se28 oryza sativ
107	42	51.2	113	2	Q6I825_OSCAG	Q6i825 oscillatori	180	41	50.0	384	2	Q7XAU4_SOYBN	Q7xau4 glyeine max
108	42	51.2	113	2	Q6I827_OSCAG	Q6i827 oscillatori	181	41	50.0	385	2	Q4B682_9BURK	Q4b682 polaromonas
109	42	51.2	113	2	Q6I831_OSCAG	Q6i831 oscillatori	182	41	50.0	396	2	Q7QNG0_ANOGA	Q7qng0 anopheles g
110	42	51.2	113	2	Q6I833_OSCAG	Q6i833 planktothri	183	41	50.0	407	2	Q3G9A3_9FIRM	Q3g9a3 syntrophomo
111	42	51.2	113	2	Q6I835_OSCAG	Q6i835 oscillatori	184	41	50.0	420	2	Q52501_BURPS	Q52501 burkholderi
112	42	51.2	118	2	Q69KQ3_ORYSA	Q69kq3 oryza sativ	185	41	50.0	455	2	Q5B577_EMENI	Q5b577 aspergillus
113	42	51.2	130	2	Q52127_9ZZZZ	Q52127 plasmid pl.	186	41	50.0	469	2	Q6XV30_9INFA	Q6xv30 influenza a
114	42	51.2	134	2	Q86445_PLARU	Q86445 planktothri	187	41	50.0	469	2	Q710U6_IACKS	Q710u6 influenza a
115	42	51.2	134	2	Q88099_3CYAN	Q88099 planktothri	188	41	50.0	474	2	Q6R7A2_9HERP	Q6r7a2 ostrleid her
116	42	51.2	134	2	Q798E8_PLARU	Q798e8 planktothri	189	41	50.0	490	2	Q7EYM6_ORYSA	Q7eym6 oryza sativ
117	42	51.2	134	2	Q798F3_3CYAN	Q798f3 planktothri	190	41	50.0	538	2	Q5Z367_NOCFA	Q5z367 nocardia fa
118	42	51.2	134	2	Q798F7_OSCAG	Q798f7 oscillatori	191	41	50.0	559	2	Q89F99_BRAJA	Q89f99 bradyrhizob
119	42	51.2	134	2	Q7B2U5_PLARU	Q7b2u5 planktothri	192	41	50.0	592	1	HPSE2_HUMAN	Q8hwq2 homo sapien
120	42	51.2	225	2	Q3FRF5_9BURK	Q3frf5 rhodoferax	193	41	50.0	606	2	Q2M1H9_HUMAN	Q2mq27 sodalis glo
121	42	51.2	231	2	Q871B0_NEUCR	Q871b0 neurospora	194	41	50.0	615	2	Q4LC08_SODGL	Q4lc08 sodalis glo
122	42	51.2	233	2	Q7R214_NEUCR	Q7r214 neurospora	195	41	50.0	606	2	Q3M7A7_ANAVT	Q3m7a7 anabaena va
123	42	51.2	266	2	Q4E222_CHRLS	Q4e222 chromohalob	196	41	50.0	615	2	Q8YRA5_ANASP	Q8yra5 anabaena sp
124	42	51.2	288	2	Q7XY15_CHLS6	Q7xy15 chlorarachn	197	41	50.0	664	2	Q2WSS4_CLOBE	Q2wss4 clostridium
125	42	51.2	318	2	Q60ZH3_CAEBR	Q60zh3 caenorhabdi	198	41	50.0	699	2	Q5W7P2_APIME	Q5w7p2 apis mellif
126	42	51.2	341	1	DRE2C_ARATH	Q8lfr2 arabidopsis	199	41	50.0	713	2	Q7YV49_CRYPV	Q7yvv49 cryptospori
127	42	51.2	343	2	Q6USL6_KLEPN	Q6usl6 klebsiella	200	41	50.0	737	2	Q58IS7_RHILO	Q58is7 rhizobium l
128	42	51.2	366	2	Q9HUS4_PSEAE	Q9hus4 pseudomonas	201	41	50.0	831	2	Q7VNB0_HAEDU	Q7vnb0 haemophilus
129	42	51.2	371	1	Y4OS_RH1SN	P55604 rhizobium s	202	41	50.0	853	2	Q86ZN0_PODAN	Q86zn0 podospora a
130	42	51.2	429	2	Q61PB3_CAEBR	Q61pb3 caenorhabdi	203	41	50.0	896	2	Q4WPA2_ASPFU	Q4wpa2 aspergillus
131	42	51.2	454	2	Q33SJO_9GAMW	Q33sj0 shewanella	204	41	50.0	1162	2	Q8D2T8_WIGBR	Q8d2t8 wigglewort
132	42	51.2	463	2	Q37TL5_SPHAR	Q37tl5 novosphingo	205	41	50.0	1220	2	Q667G7_YERPE	Q667g7 yersinia ps
133	42	51.2	493	2	Q7PH57_ANOGA	Q7ph57 anopheles g	206	41	50.0	1220	2	Q8ZH88_YERPE	Q8zh88 yersinia pe
134	42	51.2	498	2	Q3SLB7_THIDA	Q3slb7 thiobacilli	207	41	50.0	1241	2	Q8CZY3_YERPE	Q8czy3 yersinia pe
135	42	51.2	515	2	Q8T108_BOMMO	Q8t108 bombyx mori	208	41	50.0	1338	2	Q4SW64_PHACH	Q4sw64 phanerocha
136	42	51.2	527	2	Q9LRC8_SCUBA	Q9lrc8 scutellaria	209	41	50.0	382	2	Q3K4G6_PSEPF	Q3k4g6 pseudomongo
137	42	51.2	530	2	Q4HY92_GIBZE	Q4hy92 gibberella	210	40.5	49.4	640	2	Q37VM1_SPHAR	Q37vm1 novosphingo
138	42	51.2	591	2	Q5NVY0_CRYNE	Q5nvyo cryptococcu	211	40.5	49.4	86	2	Q2P9E5_XANOR	Q2p9e5 xanthomonas
139	42	51.2	591	2	Q5KEH9_CRYNE	Q5keh9 cryptococcu	212	40	48.8	86	2	Q2Q2Z1_LEPIC	Q2q2z1 leptospira
140	42	51.2	661	2	Q7PQJ2_ANOGA	Q7pqj2 anopheles g	213	40	48.8	114	2	Q8F4U7_LEPIC	Q8f4u7 leptospira
141	42	51.2	775	2	Q66593_AQUAE	Q66593 aquifex aeo	214	40	48.8	122	2	Q8YIP8_BRUME	Q8yip8 brucella me
142	42	51.2	809	2	Q53F43_HUMAN	Q53f43 homo sapien	215	40	48.8	131	2	Q80027_YEAST	Q80027 saccharomyc
143	42	51.2	809	2	Q96JS3_HUMAN	Q96js3 homo sapien	216	40	48.8	151	2	Q8KKU5_RHIET	Q8kku5 rhizobium e
144	42	51.2	833	1	M4K1_HUMAN	Q92918 homo sapien	217	40	48.8	167	2	Q304V6_CAPAN	Q304v6 capsicum an
145	42	51.2	890	2	Q8EWR8_MYCPE	Q8ewr8 mycoplasma	218	40	48.8	173	2	Q89IV6_BRAJA	Q89iv6 bradyrhizob
146	42	51.2	921	2	Q6CF18_YARLI	Q6cf18 yarrowia li	219	40	48.8	174	2	Q3RZX1_CALME	Q3rrzx1 caisticella m
147	42	51.2	1041	2	Q4QFB8_LEIMA	Q4qfb8 leishmania	220	40	48.8	178	2	Q2Z1B1_CALSA	Q2z1b1 caldwellia
148	42	51.2	1050	2	Q2IP55_9DELT	Q2ip55 anaeromyxob	221	40	48.8	196	2	Q3IHX6_PSEHT	Q3ihx6 pseudoalter
149	42	51.2	1083	2	Q5U045_MIMIV	Q5uq45 mimivirus	222	40	48.8	199	2	Q4T222_TETNG	Q4t222 tetraodon n
150	42	51.2	1366	1	CO1A2_HUMAN	P08123 homo sapien	223	40	48.8	200	1	ERF9_ARATH	Q9fe67 arabidopsis
151	42	51.2	1366	2	Q15177_HUMAN	Q15177 homo sapien	224	40	48.8	209	2	Q8ELK0_OCEIH	Q8elk0 oceanobacil
152	42	51.2	1366	2	Q7Z5S6_HUMAN	Q7z5s6 homo sapien	225	40	48.8	210	2	Q84LQ6_LYCES	Q84lq6 lycopersico
153	42	51.2	1481	2	Q5VWL1_HUMAN	Q5vwl1 homo sapien	226	40	48.8	216	2	Q2N5X0_9SPHN	Q2n5x0 erythrobact
154	42	51.2	1617	2	Q4WXY0_ASPFU	Q4wxy0 aspergillus	227	40	48.8	218	2	Q212S8_MEDTR	Q212s8 medicago tr
155	41.5	50.6	95	2	Q9IA90_ORYLA	Q9ia90 oryzias lat	228	40	48.8	218	2	Q3AIK7_SYNSC	Q3aik7 synchococc
156	41.5	50.6	220	2	Q2NXT6_XANOR	Q2nxt6 xanthomonas	229	40	48.8	218	2	Q7V7D5_PROMM	Q7v7d5 prochloroco
157	41.5	50.6	290	2	Q5GQH6_XANOR	Q5guh6 xanthomonas	230	40	48.8	220	2	Q6TKQ3_VITAE	Q6tkq3 vitis aesti
158	41.5	50.6	774	2	Q4SGZ3_TETNG	Q4sgz3 tetraodon n	231	40	48.8	222	1	ERF4_ARATH	Q80340 arabidopsis
159	41	50.0	177	2	Q5CE78_CRYHO	Q5ce78 cryptospori	232	40	48.8	222	2	Q3L0Q8_GOSHI	Q3l0q8 gossypium h
160	41	50.0	204	2	Q6RUR2_CAPAN	Q6rur2 capsicum an	233	40	48.8	222	2	Q53XI2_ARATH	Q53xi2 arabidopsis
161	41	50.0	211	2	Q32W75_CAPAN	Q32w75 capsicum an	234	40	48.8	222	2	Q84XB1_LYCES	Q84xb1 lycopersico
162	41	50.0	225	1	ERF4_TOBAC	Q40477 nicotiana t	235	40	48.8	226	2	Q41W90_DESHA	Q41w90 desulfitoba
163	41	50.0	227	1	ERF4_NICSY	Q91w49 nicotiana s	236	40	48.8	228	2	Q9WH76_9RHAB	Q9wh76 chandipura
164	41	50.0	249	2	Q6TKQ4_VITAE	Q6tkq4 vitis aesti	237	40	48.8	239	1	SFSA_AGR75	P58429 agrobacteri
165	41	50.0	252	2	Q2RYW8_SPHI	Q2ryw8 salinibacte	238	40	48.8	239	1	Q9HTL8_PSEAE	Q9htl8 pseudomonas
166	41	50.0	259	2	Q5IWL7_TOBAC	Q5iwl7 nicotiana t	239	40	48.8	243	2	Q3JB84_NITOC	Q3jb84 nitrosococc
167	41	50.0	273	2	Q5TIL3_FUGRU	Q5til3 fugu rubrip	240	40	48.8	248	2	Q8MMN3_PSEPK	Q8mmn3 pseudomonas
168	41	50.0	280	2	Q5OJ61_FUGRU	Q5oj61 fugu rubrip	241	40	48.8	250	2	Q2NMN3_9SPHN	Q2nmn3 erythrobact
169	41	50.0	306	2	Q8UCX1_AGR75	Q8ucx1 agrobacteri	242	40	48.8	252	2	Q2RXD3_RHOUR	Q2rxd3 rhodospiril
170	41	50.0	308	2	Q41BR1_GIBBEZ	Q41br1 gibberella	243	40	48.8	257	2	Q57BP7_BRUAB	Q57bp7 brucella ab
171	41	50.0	312	2	Q94HF2_ORYSA	Q94hf2 oryza sativ	244	40	48.8	257	2	Q8FZ65_BRUSU	Q8fz65 brucella su
172	41	50.0	326	2	Q7XEH8_ORYSA	Q7xbh8 oryza sativ	245	40	48.8	257	2	Q4YQB9_BRUA2	Q4yqb9 brucella ab
173	41	50.0	326	2	Q8VXC3_ORYSA	Q8vxc3 oryza sativ	246	40	48.8	258	2	Q4PKD2_9ROSI	Q4pkd2 jatrophia ch
174	41	50.0	332	2	Q84LJ7_WHEAT	Q84lj7 triticum ae	247	40	48.8	276	1	ECK2_AERPE	Q9yc05 aeropyrum p
175	41	50.0	349	2	Q8HOK1_WHEAT	Q8hok1 triticum ae	248	40	48.8	282	1	ERF5_NICSY	Q91w48 nicotiana s
176	41	50.0	355	2	Q2TN80_WHEAT	Q2tn80 triticum ae	249	40	48.8				
177	41	50.0	355	2	Q7XY26_WHEAT	Q7xy26 triticum ae	250	40	48.8				

251	40	48.8	291	1	ERF5_TOBAC	Q40478 nicotiana t	324	39.5	48.2	383	2	Q8NLE1_CORGL	Q8nle1 corynebacte
252	40	48.8	293	2	Q52242_92ZZZ	Q52242 plasmid pvt	325	39.5	48.2	387	2	Q8FLN3_COREF	Q8fln3 corynebacte
253	40	48.8	293	2	Q7W283_BORPA	Q7W282 bordetella	326	39.5	48.2	568	2	Q5B7S5_EMENI	Q5b7s5 aspergillus
254	40	48.8	293	2	Q7WR50_BORBR	Q7WR50 bordetella	327	39.5	48.2	810	2	Q82MJ1_STRAW	Q82mj1 streptomyce
255	40	48.8	313	2	Q4SHB8_TETNG	Q4shb8 tetradon n	328	39.5	48.2	2257	2	Q6K777_ORYSA	Q6k777 oryza sativ
256	40	48.8	318	2	Q6Z3H9_ORYSA	Q6z3h9 oryza sativ	329	39	47.6	46	2	Q9H3V6_HUMAN	Q9h3v6 homo sapien
257	40	48.8	319	2	Q6K4M2_STRCO	Q6k4m2 streptomyce	330	39	47.6	78	2	Q30WK2_DESDG	Q30wk2 desulfovibr
258	40	48.8	323	2	Q7NDQ9_GLOVI	Q7ndq9 gloeobacter	331	39	47.6	87	2	Q32HM4_SHIDS	Q32hm4 shigella dy
259	40	48.8	324	2	Q8LML3_ORYSA	Q8lml3 oryza sativ	332	39	47.6	93	2	Q2NXL1_XANOR	Q2nxl1 xanthomonas
260	40	48.8	326	2	Q65844_MEDSA	Q65844 medicago sa	333	39	47.6	93	2	Q5GUAL_XANOR	Q5gual xanthomonas
261	40	48.8	331	2	Q2PC80_STRAH	Q2pc80 streptomyce	334	39	47.6	95	2	Q2LQ46_9DEL1	Q2lq46 syntrophus
262	40	48.8	333	2	Q41PL9_METBU	Q41pl9 methanococc	335	39	47.6	95	2	Q6NJ54_CORDI	Q6nj54 corynebacte
263	40	48.8	341	2	Q6RZW7_VITAE	Q6rzw7 vitis aesti	336	39	47.6	96	2	Q6D9H8_ERWCT	Q6d9h8 erwina car
264	40	48.8	366	2	Q4H7W0_9DEIO	Q4h7w0 deinococcus	337	39	47.6	110	2	Q8IQ66_DROME	Q8iq66 drosophila
265	40	48.8	369	2	Q64LH6_CAPAN	Q64lh6 capsicum an	338	39	47.6	115	2	Q8UG92_AGRF5	Q8ug92 agrobacteri
266	40	48.8	369	2	Q99CE1_BRAJA	Q99cel bradyrhizob	339	39	47.6	128	2	Q6MMK9_ORYSA	Q6mmk9 oryza sativ
267	40	48.8	370	2	Q4IXG6_AZOVI	Q4ixg6 azotobacter	340	39	47.6	158	2	Q8IGF2_DROPHI	Q8igf2 drosophila
268	40	48.8	372	2	Q8LGR9_LYCES	Q8lgr9 lycopersico	341	39	47.6	158	2	Q3F5W4_9BURK	Q3f5w4 burkholderi
269	40	48.8	375	2	Q829I9_STRAW	Q829i9 streptomyce	342	39	47.6	166	2	Q7V1E5_ORYSA	Q7v1e5 oryza sativ
270	40	48.8	376	2	Q4T7U2_TETNG	Q4t7u2 tetradon n	343	39	47.6	175	2	Q7PWN1_ANOGA	Q7pwn1 anopheles g
271	40	48.8	380	2	Q4H4G2_BACCI	Q4h4g2 bacillus ci	344	39	47.6	182	2	Q7CZV8_AGRF5	Q7czv8 agrobacteri
272	40	48.8	383	2	Q2U2B8_ASPOR	Q2u2b8 aspergillus	345	39	47.6	184	2	Q9LQ28_ARATH	Q9lq28 arabidopsis
273	40	48.8	386	2	Q38068_BACTIOP	Q38068 bacterioph	346	39	47.6	192	2	Q2WB59_MAGSA	Q2wb59 magnetospi
274	40	48.8	390	2	Q8GCB3_BACCI	Q8gcb3 bacillus ci	347	39	47.6	192	2	Q9AJL8_HYDTH	Q9ajl8 hydrogenoba
275	40	48.8	391	2	Q8DJ64_SYNEL	Q8dj64 synectococc	348	39	47.6	192	2	Q2ISL7_RHOPA	Q2isl7 rhodopseu
276	40	48.8	391	2	Q9RJP0_STRCO	Q9rjp0 streptomyce	349	39	47.6	195	2	Q8S2S7_THEHA	Q8s2s7 thellungiel
277	40	48.8	398	2	Q5FTR6_GLUOX	Q5ftr6 gluconobact	350	39	47.6	196	1	WBBJ_ECOLI	Wbj750 escherichia
278	40	48.8	410	2	Q3UC89_MOUSE	Q3uc89 m bone marr	351	39	47.6	197	2	Q6ZAH8_ORYSA	Q6zah8 oryza sativ
279	40	48.8	426	2	Q2P244_XANOR	Q2p244 xanthomonas	352	39	47.6	198	2	Q6L4M2_ORYSA	Q6l4m2 oryza sativ
280	40	48.8	426	2	Q3BS70_XANC5	Q3bs70 xanthomonas	353	39	47.6	203	2	Q3FQ51_9BURK	Q3fq51 rhodofera
281	40	48.8	426	2	Q4UVU6_XANC8	Q4uvu6 xanthomonas	354	39	47.6	207	2	Q4R4W9_MACFA	Q4r4w9 macaca fasc
282	40	48.8	426	2	Q8P891_XANCP	Q8p891 xanthomonas	355	39	47.6	212	2	Q7VT35_BORPE	Q7vt35 bordetella
283	40	48.8	431	2	Q3TDG5_MOUSE	Q3tdg5 mus musculu	356	39	47.6	212	2	Q7WFI2_BORBR	Q7wfi2 bordetella
284	40	48.8	442	2	Q8PJG6_XANAC	Q8pjg6 xanthomonas	357	39	47.6	221	2	Q98545_PBCV1	Q98545 paramecium
285	40	48.8	449	2	Q2LFX4_9INFA	Q2lfx4 influenza a	358	39	47.6	231	2	Q4W6U0_NICBE	Q4w6u0 nicotiana b
286	40	48.8	449	2	Q2LFX5_9INFA	Q2lfx5 influenza a	359	39	47.6	237	2	Q9FR02_TOBAC	Q9fr02 nicotiana t
287	40	48.8	455	2	Q747L5_GEOSL	Q747l5 geobacter s	360	39	47.6	238	2	Q9ZR85_STYHA	Q9zr85 stylosanthe
288	40	48.8	493	2	Q8SYJ5_DROME	Q8syj5 drosophila	361	39	47.6	239	2	Q9ZR83_STYHA	Q9zr83 stylosanthe
289	40	48.8	493	2	Q9VMN1_DROME	Q9vmn1 drosophila	362	39	47.6	240	2	Q8H6S9_LYCES	Q8h6s9 lycopersico
290	40	48.8	496	2	Q2SGJ8_9GAWM	Q2sgj8 habella che	363	39	47.6	245	2	Q8GZE9_LYCES	Q8gze9 lycopersico
291	40	48.8	511	2	Q5GZ29_XANOR	Q5gz29 xanthomonas	364	39	47.6	246	2	Q8CCT9_MOUSE	Q8cct9 mus musculu
292	40	48.8	543	2	Q7MDK8_VIBVY	Q7mdk8 vibrio vuln	365	39	47.6	247	2	Q6PHC5_BRARE	Q6phc5 brachydanio
293	40	48.8	543	2	Q8D6Q1_VIBVU	Q8d6q1 vibrio vuln	366	39	47.6	255	2	Q852K1_ENTFA	Q852k1 enterococcu
294	40	48.8	599	2	Q5HMK3_STAEQ	Q5hmk3 staphylococ	367	39	47.6	259	2	Q47L96_THPEY	Q47l96 themobifid
295	40	48.8	629	2	Q3VGF9_9SPHN	Q3vgf9 sphingopyxi	368	39	47.6	273	2	Q7X649_ORYSA	Q7x649 oryza sativ
296	40	48.8	668	2	Q6PG56_MOUSE	Q6pg56 mus musculu	369	39	47.6	273	2	Q9AQU3_ORYSA	Q9aqu3 oryza sativ
297	40	48.8	736	2	Q4QF12_LEIMA	Q4qf12 leishmania	370	39	47.6	277	2	Q5LQ59_SILPO	Q5lq59 silicibacte
298	40	48.8	773	2	Q3F8B8_9RHO	Q3f8b8 azoarcus ev	371	39	47.6	278	2	Q6RZM8_VITAE	Q6rzm8 vitis aesti
299	40	48.8	822	2	Q33FC6_9RHO	Q33fc6 azoarcus ev	372	39	47.6	279	2	Q4K6B4_PSEPF	Q4k6b4 pseudomonas
300	40	48.8	822	2	Q83HG9_TROW8	Q83hg9 tropheryma	373	39	47.6	279	2	Q4K5I0_PSEPF	Q4k5i0 pseudomonas
301	40	48.8	822	2	Q83GT4_TROWT	Q83gt4 tropheryma	374	39	47.6	279	2	Q4K5I0_PSEPF	Q4k5i0 pseudomonas
302	40	48.8	879	2	Q7TPK0_RAT	Q7tpk0 rattus norv	375	39	47.6	280	2	Q4BDN5_PSE14	Q4bdn5 pseudomonas
303	40	48.8	893	1	RUSC1_MOUSE	Q8b926 mus musculu	376	39	47.6	280	2	Q4ZNR4_PSEU2	Q4zn4 pseudomonas
304	40	48.8	958	2	Q49931_MYCLE	Q49931 mycobacteri	377	39	47.6	280	2	Q6AC01_LEIXX	Q6ac01 leifsonia x
305	40	48.8	979	2	Q73TG7_MYCPA	Q73tg7 mycobacteri	378	39	47.6	280	2	Q87VX8_PSESM	Q87vx8 pseudomonas
306	40	48.8	988	2	Q6CPA2_KULUA	Q6cpa2 kluyveromyc	379	39	47.6	283	2	Q2UIT8_ASPOR	Q2uit8 aspergillus
307	40	48.8	1014	2	Q9MAR9_ARATH	Q9mar9 arabidopsis	380	39	47.6	290	2	Q3QN11_9RHOB	Q3qn11 silicibacte
308	40	48.8	1031	2	Q2JB15_9ACTO	Q2jb15 frankia sp.	381	39	47.6	291	2	Q3W3F0_9ACTO	Q3w3f0 frankia sp.
309	40	48.8	1126	1	VIA_AMVLE	P03589 alfalfa mos	382	39	47.6	292	2	Q441N3_SOLUS	Q441n3 solibacter
310	40	48.8	1191	2	Q9RHV0_STRRO	Q9rhv0 streptomyce	383	39	47.6	292	2	Q471K0_RALEBJ	Q471k0 ralsconia e
311	40	48.8	1440	2	Q6VY42_9CAUD	Q6vy42 bacterioph	384	39	47.6	293	2	Q2J3E3_RHOPA	Q2j3e3 rhodopseu
312	40	48.8	1445	2	Q93251_RANCA	Q93251 rana cateb	385	39	47.6	296	2	Q5TW01_ANOGR	Q5tw01 anopheles g
313	40	48.8	1450	2	Q9Y1B4_CNPY	Q9y1b4 cynops pyrr	386	39	47.6	298	2	Q3HBV9_TRIER	Q3hbv9 trichodesmi
314	40	48.8	1666	2	Q8LP68_CHLRE	Q8lp68 chlamydomon	387	39	47.6	308	1	YDDO_ECOLI	Yddo22 escherichia
315	40	48.8	1698	2	Q94438_CHIPA	Q94438 chironomus	388	39	47.6	308	2	Q5TW02_ANOGA	Q5tw02 anopheles g
316	40	48.8	1871	2	Q9Z5K6_MYCLE	Q9z5k6 mycobacteri	389	39	47.6	308	2	Q320S2_SHIBS	Q320s2 shigella bo
317	40	48.8	2152	2	Q9ALM5_9PSEU	Q9alm5 saccharopol	390	39	47.6	308	2	Q32G18_SHIDS	Q32g18 shigella dy
318	40	48.8	3295	2	Q83X71_STRRO	Q83x71 streptomyce	391	39	47.6	308	2	Q8X4Z6_ECO57	Q8x4z6 escherichia
319	40	48.8	3651	2	Q83X69_STRRO	Q83x69 streptomyce	392	39	47.6	310	2	Q826M4_STRAW	Q826m4 streptomyce
320	40	48.8	4928	2	Q9ALM3_9PSEU	Q9alm3 saccharopol	393	39	47.6	312	2	Q4C4C0_CROWT	Q4c4c0 crocospaer
321	40	48.8	5588	2	Q9ALM2_9PSEU	Q9alm2 saccharopol	394	39	47.6	313	2	Q2X1M4_PSEPU	Q2x1m4 pseudomona
322	39.5	48.2	124	2	Q4J8I6_SULAC	Q4j8i6 sulfolobus	395	39	47.6	316	2	Q37S31_SPHAR	Q37s31 novosphingo
323	39.5	48.2	131	1	HIS3_PYRAE	Q8zy39 pyrobaculum	396	39	47.6	318	2	Q7XSD2_ORYSA	Q7xsd2 oryza sativ

397	39	47.6	319	2	Q6ABJ4_PROAC	Q6abj4 propionibac	470	39	47.6	572	2	Q3B2X5_PELLD	Q3b2x5 pelodictyon
398	39	47.6	322	2	Q4J9E0_SULAC	Q4j9e0 sulfolobus	471	39	47.6	581	2	Q4I1M5_GIBZE	Q4i1m5 gibberella
399	39	47.6	322	2	Q4KB39_PSEF5	Q4kb39 pseudomonas	472	39	47.6	594	2	Q4HVM2_GIBZE	Q4hvm2 gibberella
400	39	47.6	326	2	Q96D47_HUMAN	Q96d47 homo sapien	473	39	47.6	609	2	Q3ZC37_BOVIN	Q3zc37 bos taurus
401	39	47.6	327	2	Q7PSR8_ANOGA	Q7psr8 anopheles g	474	39	47.6	617	2	Q3BGC6_ACTSC	Q3bgc6 actinobacil
402	39	47.6	327	2	Q6TXK7_LYCES	Q6txk7 lycopersico	475	39	47.6	621	2	Q41CK5_9BACI	Q41ck5 exigubacte
403	39	47.6	327	2	Q2XDC6_PSEPU	Q2xdc6 pseudomonas	476	39	47.6	622	2	Q6CGF8_YARLI	Q6cgf8 yarrowia l1
404	39	47.6	328	2	Q7XJ72_ORISA	Q7xj72 oryza sativ	477	39	47.6	669	2	Q982K1_RHILO	Q982k1 rhizobium l1
405	39	47.6	328	2	Q7XAD5_ORISA	Q7xad5 oryza sativ	478	39	47.6	674	2	Q6BSC1_DEBHA	Q6bsc1 debaryomyce
406	39	47.6	330	2	Q9F7C9_RRHIZ	Q9f7c9 agrobacteri	479	39	47.6	687	2	Q3ZDP9_ARAGE	Q3zdp9 arabis gemm
407	39	47.6	343	2	Q8CFE0_MOUSE	Q8cef0 mus musculu	480	39	47.6	693	2	Q3SM98_THIDA	Q3sm98 thiobacillu
408	39	47.6	365	1	CATD_SHEEP	Q9mz58 ovis aries	481	39	47.6	703	2	Q8RXN0_ARATH	Q8rxn0 arabidopsis
409	39	47.6	365	1	GAL7_YEAST	P08431 saccharomyc	482	39	47.6	713	2	Q3PB31_PARDE	Q3pb31 paracoccu
410	39	47.6	365	2	Q3B129_PELLD	Q3b129 pelodictyon	483	39	47.6	744	1	NSF_CRIGR	P18708 cricetulus
411	39	47.6	366	2	Q3XP07_PPROT	Q3xpf07 magnetococc	484	39	47.6	744	1	NSF_HUMAN	P46459 homo sapien
412	39	47.6	378	2	Q4RLR4_TETNG	Q4rlr4 tetraodon n	485	39	47.6	744	1	NSF_MOUSE	P46460 mus musculu
413	39	47.6	379	1	FTSZ_MYCBO	P64171 mycobacteri	486	39	47.6	744	2	Q8N6D7_HUMAN	Q8n6d7 homo sapien
414	39	47.6	379	1	FTSZ_MYCBO	P64170 mycobacteri	487	39	47.6	744	2	Q8N6D7_HUMAN	Q8n6d7 homo sapien
415	39	47.6	382	1	SUBT_BACAM	P00782 bacillus am	488	39	47.6	744	2	Q8C3R2_MOUSE	Q8c3r2 mus musculu
416	39	47.6	382	1	Q5NLV2_ZYMMO	Q5nlv2 zymomonas m	489	39	47.6	744	2	Q8C3R2_MOUSE	Q8c3r2 mus musculu
417	39	47.6	386	2	Q5AYL0_EMENI	Q5ayl0 aspergillus	490	39	47.6	744	2	Q923C6_MOUSE	Q923c6 m n-ethylma
418	39	47.6	388	2	Q3GVA9_BACTO	Q3gv49 nocardioid	491	39	47.6	744	2	Q9QUL6_RAT	Q9qul6 rattus norv
419	39	47.6	389	2	Q44H95_CHRSL	Q44h95 chromohalob	492	39	47.6	787	2	Q2VAD3_HPBUD	Q2vad3 duck hepati
420	39	47.6	390	2	Q8RD18_THETN	Q8rdl8 thermoaer	493	39	47.6	804	2	Q3ITQ0_NATPD	Q3itq0 natronomona
421	39	47.6	406	2	Q3FN14_9BURK	Q3fn14 rhodoferrax	494	39	47.6	812	2	Q4WG06_ASPFU	Q4wg06 aspergillus
422	39	47.6	407	2	Q46F29_METBA	Q46f29 methanosarc	495	39	47.6	821	2	Q2ULM7_ASPOR	Q2ulm7 aspergillus
423	39	47.6	410	2	Q9CQJ5_MYCLE	Q9ccj5 mycobacteri	496	39	47.6	828	2	Q6S503_ARATH	Q6s503 arabidopsis
424	39	47.6	420	2	Q7U360_BORPE	Q7u360 bordetella	497	39	47.6	865	2	Q5GH14_AERPU	Q5gh14 aeromonas p
425	39	47.6	420	2	Q7U371_BORPA	Q7u371 bordetella	498	39	47.6	866	2	Q9L5D5_AERHY	Q9l5d5 aeromonas h
426	39	47.6	420	2	Q7U382_BORBR	Q7u382 bordetella	499	39	47.6	867	2	Q36LU2_MARHY	Q36lu2 clostridiate
427	39	47.6	424	2	Q65CX5_BACLD	Q65cx5 bacillus li	500	39	47.6	869	2	Q2WLJ0_CLOBE	Q2wlj0 clostridium
428	39	47.6	425	2	Q62NF0_BACLD	Q62nf0 bacillus li	501	39	47.6	888	2	Q4WMU3_ASPFU	Q4wmu3 aspergillus
429	39	47.6	426	2	Q8KCT1_CHLTE	Q8kct1 chlorobium	502	39	47.6	891	2	Q54RN3_DICDI	Q54rn3 dictyosteli
430	39	47.6	427	2	Q58CY1_BOVIN	Q58cy1 bos taurus	503	39	47.6	923	2	Q4SBM6_TETNG	Q4sbm6 tetraodon n
431	39	47.6	429	1	YVFO_BACSU	Q07013 bacillus su	504	39	47.6	939	2	Q2RYX4_9SPHI	Q2ryx4 salinibacte
432	39	47.6	442	2	Q6N7G8_RHOPA	Q6n7g8 rhodopseudo	505	39	47.6	951	2	Q6FK34_CANGA	Q6fk34 candida gla
433	39	47.6	442	2	Q37N81_RHOPA	Q37n81 rhodopseudo	506	39	47.6	1023	1	DPOL_ADEB3	Q3tp88 mus musculu
434	39	47.6	450	2	Q4IDP2_GIBZE	Q4idd2 gibberella	507	39	47.6	1048	2	Q374A5_RHOSA	Q374a5 bovine aden
435	39	47.6	460	2	Q6DX87_9INFA	Q6dx87 influenza a	508	39	47.6	1050	2	Q4X0U9_ASPFU	Q4x0u9 aspergillus
436	39	47.6	461	2	Q6DMW3_9INFA	Q6dmw3 influenza a	509	39	47.6	1097	2	Q870T4_NEUCR	Q870t4 neurospora
437	39	47.6	462	2	Q6DW21_9INFA	Q6dw21 influenza a	510	39	47.6	1098	2	Q43VM3_SOLUS	Q43vm3 solibacter
438	39	47.6	464	2	Q5DL12_9INFA	Q5dl12 influenza a	511	39	47.6	1124	2	Q5KBB6_CRYNE	Q5kbb6 cryptococcu
439	39	47.6	466	2	Q6DMW1_9INFA	Q6dmw1 influenza a	512	39	47.6	1134	2	Q6IRN3_XENLA	Q6irn3 xenopus lae
440	39	47.6	466	2	Q6DMW7_9INFA	Q6dmw7 influenza a	513	39	47.6	1189	2	Q2VB19_CHICK	Q2vb19 gallus gall
441	39	47.6	467	2	Q6DX55_9INFA	Q6dx55 influenza a	514	39	47.6	1190	2	Q3R784_XYLFA	Q3r784 xyella fas
442	39	47.6	469	2	Q98077_9INFA	Q98077 influenza a	515	39	47.6	1190	2	Q87DV4_XYLFA	Q87dv4 xyella fas
443	39	47.6	470	1	NRAM_IADM2	Q07573 influenza a	516	39	47.6	1372	1	COLA2_MOUSE	Q07573 influenza a
444	39	47.6	470	1	NRAM_IADM2	Q07577 influenza a	517	39	47.6	1372	1	COLA2_MOUSE	Q07577 influenza a
445	39	47.6	470	1	NRAM_IAMAE	Q07583 influenza a	518	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
446	39	47.6	470	1	NRAM_IAMAE	Q07584 influenza a	519	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
447	39	47.6	470	1	NRAM_IAMAE	Q07585 influenza a	520	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
448	39	47.6	470	2	Q595Z2_IATKC	Q595z2 influenza a	521	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
449	39	47.6	470	2	Q5UG07_9INFA	Q5ug07 influenza a	522	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
450	39	47.6	470	2	Q6DMW1_9INFA	Q6dmw1 influenza a	523	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
451	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	524	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
452	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	525	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
453	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	526	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
454	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	527	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
455	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	528	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
456	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	529	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
457	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	530	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
458	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	531	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
459	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	532	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
460	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	533	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
461	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	534	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
462	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	535	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
463	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	536	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
464	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	537	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
465	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	538	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
466	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	539	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
467	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	540	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
468	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	541	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
469	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	542	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu

543	38.5	47.0	314	2	Q743T7_MYCPA	Q743T7	mycobacteri	616	38	46.3	244	2	Q6Q4I4_LYCES	O6q414 lycopersico
544	38.5	47.0	382	2	Q2X991_PSEPU	Q2X991	pseudomonas	617	38	46.3	246	2	Q2J143_ARATH	O2j143 arabidopsis
545	38.5	47.0	382	2	Q8C69_PSEPK	Q8c69	pseudomonas	618	38	46.3	248	1	AP23_ARATH	F42736 arabidopsis
546	38.5	47.0	442	2	Q7F28_9INFA	Q7f28	influenza a	619	38	46.3	248	1	PT16_LYCES	O04682 lycopersico
547	38.5	47.0	561	2	Q5Q77_CRYNE	Q5q77	cryptococcus	620	38	46.3	249	2	Q67MY7_SYMTH	O67my7 symbiobacte
548	38.5	47.0	561	2	Q5KF9_CRYNE	Q5kfy9	cryptococcus	621	38	46.3	251	2	O65242_TOBAC	O65242 nicotiana t
549	38.5	47.0	735	2	Q5C8D8_9THEM	Q5cbd8	thermotoga	622	38	46.3	252	2	O2HY03_PENCA	O2hy03 penicillium
550	38	46.3	84	2	Q98N19_RHILO	Q98n19	rhizobium l	623	38	46.3	254	2	Q479V4_DECAR	Q479v4 dechloromon
551	38	46.3	87	2	Q48933_9POTV	Q48933	plum pox vi	624	38	46.3	258	2	Q332T4_RHOFA	Q332t4 rhodopsendo
552	38	46.3	99	2	Q48BV2_PSE14	Q48bv2	pseudomonas	625	38	46.3	260	2	Q84XB2_LYCES	Q84xb2 lycopersico
553	38	46.3	101	2	Q6PDS4_MOUSE	Q6pds4	mus musculus	626	38	46.3	260	2	Q6RJ36_LYCES	Q6rj36 lycopersico
554	38	46.3	102	2	Q8WY95_HUMAN	Q8wy95	homo sapien	627	38	46.3	263	2	Q92NLO_RHIME	Q92nlo rhizobium m
555	38	46.3	106	2	Q95M02_HORSE	Q95m02	equus caball	628	38	46.3	264	2	Q6QSS5_CAPAN	O6qqss5 capsicum an
556	38	46.3	118	2	Q6K8B4_ORYSA	Q6k8b4	oryza sativ	629	38	46.3	270	2	Q3W647_9ACTO	Q3w647 frankia sp.
557	38	46.3	124	2	Q7XD22_ORYSA	Q7xd22	oryza sativ	630	38	46.3	272	2	Q32W72_CAPAN	Q32w72 capsicum an
558	38	46.3	128	2	Q6J988_ARATH	Q6j988	arabidopsis	631	38	46.3	272	2	Q2S1U8_9SPHI	Q2s1u8 salinibacte
559	38	46.3	131	2	Q480S2_COLP3	Q480s2	colwellia p	632	38	46.3	274	2	Q5DZRI_VIBF1	Q5dzt1 vibrio fisc
560	38	46.3	133	2	Q455G9_9BURK	Q455g9	burkholderi	633	38	46.3	275	1	ECK2_SULTO	O97509 sulfolobus
561	38	46.3	133	2	Q4LLG1_9BURK	Q4llg1	burkholderi	634	38	46.3	276	2	Q4JB28_SULAC	Q4jb28 sulfolobus
562	38	46.3	136	2	Q8X8X1_STRCO	Q8x8x1	streptomyce	635	38	46.3	276	2	Q32W74_CAPAN	Q32w74 capsicum an
563	38	46.3	139	2	Q9LTC6_ARATH	Q9ltc6	arabidopsis	636	38	46.3	276	2	Q9CY15_MOUSE	O9cy15 mus musculu
564	38	46.3	141	2	Q6H7R2_ORYSA	Q6h7r2	oryza sativ	637	38	46.3	277	2	Q394Q7_BURS3	Q394q7 burkholderi
565	38	46.3	141	2	Q8H075_ORYSA	Q8h075	oryza sativ	638	38	46.3	279	2	Q2IKK1_9DELT	O2ikk1 anaeromyxob
566	38	46.3	144	2	Q96NC1_HUMAN	Q96nc1	homo sapien	639	38	46.3	280	2	O81365_PRUAR	O81365 prunus arme
567	38	46.3	144	2	Q4T622_TETNG	Q4t622	tetraodon n	640	38	46.3	282	1	ERP6_ARATH	O8v231 arabidopsis
568	38	46.3	146	2	Q8LI66_ORYSA	Q8li66	oryza sativ	641	38	46.3	289	2	Q37QC5_SPHAR	Q37qcs novosphingo
569	38	46.3	147	2	Q5AUL9_EMENI	Q5aul9	aspergillus	642	38	46.3	292	2	Q5IR87_HUMAN	O5ir87 homo sapien
570	38	46.3	154	2	Q95M03_HORSE	Q95m03	equus caball	643	38	46.3	292	2	Q67RC7_SYMTH	O67rc7 symbiobacte
571	38	46.3	156	2	Q4J7B6_SULAC	Q4j7b6	sulfolobus	644	38	46.3	293	2	Q4ASD3_9BURK	O4aes3 polaromonas
572	38	46.3	162	2	Q5VMU7_ORYSA	Q5vmu7	oryza sativ	645	38	46.3	297	2	Q4NFR1_9MICC	O4nfr1 arthrobacte
573	38	46.3	162	2	Q8SB09_ORYSA	Q8sb09	oryza sativ	646	38	46.3	297	2	Q660I2_BORBA	Q660i2 borrelia ga
574	38	46.3	165	2	Q75UJ5_CUCME	Q75uj5	cucumis mel	647	38	46.3	297	2	O51631_BORBU	O51631 borrelia bu
575	38	46.3	171	2	Q3XUN0_9PROT	Q3xun0	magnetococc	648	38	46.3	298	2	Q2SKB4_9GAMM	O2skb4 habella che
576	38	46.3	171	2	Q84B14_VIBF1	Q84b14	vibrio fisc	649	38	46.3	300	2	Q33MB5_METHU	Q33mb5 methanospir
577	38	46.3	172	2	Q7X1L1_ORYSA	Q7x1l1	oryza sativ	650	38	46.3	302	2	O6W1U6_RHISN	O6w1u6 rhizobium s
578	38	46.3	173	2	Q5Y4C5_ARATH	Q5y4c5	arabidopsis	651	38	46.3	302	2	Q5Z2I4_NOCPA	O5z2i4 nocardia fa
579	38	46.3	182	2	Q3AX61_SYNS9	Q3ax61	synecococc	652	38	46.3	309	2	Q7BU48_STRAM	Q7bu48 streptomyce
580	38	46.3	183	2	Q7BTW3_YERPE	Q7btw3	yersinia pe	653	38	46.3	314	2	Q973B5_SULTO	O973b5 sulfolobus
581	38	46.3	183	2	Q2KWZ7_BORAV	Q2kwz7	bordetella	654	38	46.3	317	2	O23113_ARATH	O23113 arabidopsis
582	38	46.3	185	2	Q7R0F2_GIALA	Q7r0f2	giardia lam	655	38	46.3	319	2	O672K6_GOSBA	O672k6 gossypium b
583	38	46.3	188	2	Q82QN4_STRAM	Q82qn4	streptomyce	656	38	46.3	321	2	Q619N9_CAEBR	O619n9 caenorhabdi
584	38	46.3	190	2	Q5GA65_LYCES	Q5ga65	lycopersico	657	38	46.3	322	2	Q8BLU3_MOUSE	Q8blu3 mus musculu
585	38	46.3	190	2	Q6AMA7_DESPS	Q6ama7	desulfotale	658	38	46.3	323	2	Q97UL6_SULSO	O97ul6 sulfolobus
586	38	46.3	192	2	Q4ELK3_TRYCR	Q4elk3	trypanosoma	659	38	46.3	324	2	Q4IFS5_GIBZE	O4ifs5 gibberella
587	38	46.3	192	2	Q3G5V8_9DELT	Q3g5v8	pelobacter	660	38	46.3	327	2	O2NB19_9SPHN	O2nb19 erythrobact
588	38	46.3	193	2	Q8S8T1_ARATH	Q8s8t1	arabidopsis	661	38	46.3	329	2	O6QX54_COCPA	O6qx54 coffea cane
589	38	46.3	194	2	Q6YVX0_ORYSA	Q6yvxo	oryza sativ	662	38	46.3	330	2	Q6SEV3_9POTV	O6sev3 plum pox vi
590	38	46.3	195	2	Q68725_YERPE	Q68725	yersinia pe	663	38	46.3	332	2	Q6ESV2_9POTV	O6esv2 plum pox vi
591	38	46.3	195	2	Q7ARG3_YERPE	Q7arg3	yersinia pe	664	38	46.3	333	2	Q6ESV1_9POTV	O6esv1 plum pox vi
592	38	46.3	195	2	Q935A9_SALT1	Q935a9	salmonella	665	38	46.3	335	1	DRE2A_ARATH	O82132 arabidopsis
593	38	46.3	196	2	Q3CRC8_ALTAT	Q3crc8	pseudolater	666	38	46.3	335	2	O693B3_ORYSA	O693b3 oryza sativ
594	38	46.3	201	2	Q53N79_ORYSA	Q53n79	oryza sativ	667	38	46.3	335	2	O6QJK9_9LAMI	O6qjk9 fraxinus pe
595	38	46.3	207	1	FRDA_MOUSE	Q35943	mus musculu	668	38	46.3	338	2	Q4D5F6_TRYCR	O4dsif6 trypanosoma
596	38	46.3	207	2	Q9KLL4_VIBCH	Q9kll4	vibrio chol	669	38	46.3	338	2	Q2LDQ6_ARAHY	O2ldq6 arachis hyp
597	38	46.3	207	2	Q3TV21_MOUSE	Q3tv21	mus musculu	670	38	46.3	340	2	O6Q9W4_9BRAS	O6q9w4 thellungiel
598	38	46.3	207	2	Q3UG34_MOUSE	Q3ug34	mus musculu	671	38	46.3	340	2	O8PPG0_XANAC	O8ppg0 xanthomonas
599	38	46.3	210	1	FRDA_HUMAN	Q16595	homo sapien	672	38	46.3	340	2	O6PW64_9INFA	O6pw64 influenza a
600	38	46.3	210	1	FRDA_MACFA	Q8hxx9	macaca fasc	673	38	46.3	340	2	Q6PW67_9INFA	O6pw67 influenza a
601	38	46.3	210	2	Q5VZ01_HUMAN	Q5vz01	homo sapien	674	38	46.3	342	2	Q7X826_ORYSA	Q7x826 oryza sativ
602	38	46.3	210	2	Q9LEME_CATRO	Q9leme	catharanthu	675	38	46.3	344	2	Q8GTB5_CICAR	O8gte5 cicier ariet
603	38	46.3	211	2	Q3YRL5_EHRCJ	Q3yrl5	ehrllichia c	676	38	46.3	345	2	Q3SE63_9BRAD	O3se63 bradyrhizob
604	38	46.3	211	2	Q5HAS8_EHRRW	Q5has8	ehrllichia r	677	38	46.3	346	2	O8JZ32_ARMV	O8jz32 arabis mosa
605	38	46.3	214	2	Q3FR33_LYCES	Q3fr33	lycopersico	678	38	46.3	358	2	O8L9V3_ARATH	O8l9v3 arabidopsis
606	38	46.3	217	2	Q5FD64_EHRRW	Q5fd64	ehrllichia r	679	38	46.3	358	2	Q9SSA8_ARATH	Q9ssa8 arabidopsis
607	38	46.3	217	2	Q5FFU6_EHRRG	Q5ffu6	ehrllichia r	680	38	46.3	362	1	QUEA_DEIRA	O9rul9 deinococcus
608	38	46.3	218	1	ERF1B_ARATH	Q8ldc8	arabidopsis	681	38	46.3	362	2	O69Y47_ORYSA	O69y47 oryza sativ
609	38	46.3	218	2	Q5YN71_NOCPA	Q5yn71	nocardia fa	682	38	46.3	362	2	O9SP16_ORYSA	O9sp16 oryza sativ
610	38	46.3	222	2	Q9AP70_9BACT	Q9ap70	uncultured	683	38	46.3	366	2	Q2JU07_9CYAN	Q2ju07 cyanobacter
611	38	46.3	222	2	Q9AP71_9BACT	Q9ap71	uncultured	684	38	46.3	374	1	RM02_YARLI	O6c9c4 yarrowia li
612	38	46.3	234	2	Q2LUS9_9DELT	Q2lus9	syntrophus	685	38	46.3	375	2	Q382R2_9TRYP	Q382r2 trypanosoma
613	38	46.3	235	2	Q3HW33_CAPAN	Q3hw33	capsicum an	686	38	46.3	375	2	O8VXK9_FAGSY	O8vxk9 fagus sylv
614	38	46.3	239	2	Q8H312_ORYSA	Q8h312	oryza sativ	687	38	46.3	378	1	YN9B_YEAST	P52923 saccharomyc
615	38	46.3	239	2	Q93A09_THIFE	Q93a09	thiobacillu	688	38	46.3	378	2	Q70AB2_FAGSY	Q70ab2 fagus sylv

689	38	46.3	380	2	Q51CN9_9ROSI	Q51cn9	populus alb	762	38	46.3	469	2	Q6XV56_9INFA	Q6xv56	influenza a
690	38	46.3	381	2	Q52QX1_MANES	Q52qx1	manihot esc	763	38	46.3	469	2	Q6XV57_9INFA	Q6xv57	influenza a
691	38	46.3	385	2	Q2PEZ6_TRIPR	Q2pez6	trifolium p	764	38	46.3	469	2	Q6XV59_9INFA	Q6xv59	influenza a
692	38	46.3	387	2	Q7Y1W2_TOBAC	Q7y1w2	nicotiana t	765	38	46.3	469	2	Q6XV60_9INFA	Q6xv60	influenza a
693	38	46.3	389	2	Q37G65_RHOPA	Q37g65	rhodopseudo	766	38	46.3	469	2	Q6XV64_9INFA	Q6xv64	influenza a
694	38	46.3	390	2	Q31OR1_GOSHI	Q31or1	gossypium h	767	38	46.3	469	2	Q6XV65_9INFA	Q6xv65	influenza a
695	38	46.3	390	2	Q34C06_RHOPA	Q34c06	rhodopseudo	768	38	46.3	469	2	Q6XV66_9INFA	Q6xv66	influenza a
696	38	46.3	390	2	Q3PS76_NITHA	Q3ps76	nitrobacter	769	38	46.3	469	2	Q6XV69_9INFA	Q6xv69	influenza a
697	38	46.3	390	2	Q3SNC2_NITWN	Q3snc2	nitrobacter	770	38	46.3	469	2	Q6XV70_9INFA	Q6xv70	influenza a
698	38	46.3	390	2	Q2J016_RHOPA	Q2j016	rhodopseudo	771	38	46.3	469	2	Q6XV72_9INFA	Q6xv72	influenza a
699	38	46.3	390	2	Q6N1U5_RHOPA	Q6n1u5	rhodopseudo	772	38	46.3	469	2	Q7TF24_9INFA	Q7tf24	influenza a
700	38	46.3	392	2	Q4JDH2_3PROT	Q4jdh2	nitrospir	773	38	46.3	469	2	Q7TF23_9INFA	Q7tf23	influenza a
701	38	46.3	394	2	Q2RPH0_RHORU	Q2rph0	rhodospirill	774	38	46.3	469	2	Q80M09_9INFA	Q80mj9	influenza a
702	38	46.3	396	2	Q3KZ43_GOSHI	Q3kz43	gossypium h	775	38	46.3	469	2	Q82555_9INFA	Q82555	influenza a
703	38	46.3	399	2	Q4HQZ6_CAMUP	Q4hqz6	campylobact	776	38	46.3	469	2	Q8B197_9INFA	Q8b197	influenza a
704	38	46.3	402	1	MURD_CAMJR	Q5hw34	campylobact	777	38	46.3	469	2	Q8B198_9INFA	Q8b198	influenza a
705	38	46.3	405	2	Q39LQ6_BURS3	Q39lq6	burkholderi	778	38	46.3	469	2	Q8JSE0_9INFA	Q8jse0	influenza a
706	38	46.3	416	1	CCA_SULAC	Q4j9a0	sulfolobus	779	38	46.3	469	2	Q91BI1_9INFA	Q91bi1	influenza a
707	38	46.3	417	2	Q6BJW5_DEBHA	Q6bjw5	debaromyce	780	38	46.3	469	2	Q91BI2_9INFA	Q91bi2	influenza a
708	38	46.3	418	2	Q93H82_STRAW	Q93h82	streptomyce	781	38	46.3	469	2	Q9WJT1_9INFA	Q9wjt1	influenza a
709	38	46.3	424	2	Q4IS30_GIBZE	Q4is30	gibberella	782	38	46.3	469	2	Q9YNL3_9INFA	Q9ynl3	influenza a
710	38	46.3	425	2	Q4WB80_ASPTU	Q4wb80	aspergillus	783	38	46.3	469	2	Q9IGQ0_IANJ	Q9igq0	influenza a
711	38	46.3	427	2	Q2LFT2_9INFA	Q2lft2	influenza a	784	38	46.3	469	2	Q2LFS5_9INFA	Q2lfs5	influenza a
712	38	46.3	430	2	Q6FUF6_CANGA	Q6fuf6	candida gla	785	38	46.3	469	2	Q2LFS7_9INFA	Q2lfs7	influenza a
713	38	46.3	432	2	Q483Y5_COLP3	Q483y5	colwellia p	786	38	46.3	469	2	Q693C5_9INFA	Q693c5	influenza a
714	38	46.3	437	1	WTM1_YEAST	Q12363	saccharomyc	787	38	46.3	469	2	Q6DQJ4_9INFA	Q6dqj4	influenza a
715	38	46.3	437	2	Q5A0B9_EMENI	Q5a0b9	aspergillus	788	38	46.3	469	2	Q6DSV3_9INFA	Q6dsv3	influenza a
716	38	46.3	437	2	Q3RT53_RALME	Q3rt53	raistonia m	789	38	46.3	469	2	Q6LZB9_9INFA	Q6lzb9	influenza a
717	38	46.3	441	2	Q6XV63_9INFA	Q6xv63	influenza a	790	38	46.3	469	2	Q89473_9INFA	Q89473	influenza a
718	38	46.3	442	2	Q2LFT3_9INFA	Q2lft3	influenza a	791	38	46.3	469	2	Q21DG4_9INFA	Q21dg4	influenza a
719	38	46.3	443	2	Q4WYV3_ASPTU	Q4wyv3	aspergillus	792	38	46.3	470	1	NRAM_IACHU	Q11485	influenza a
720	38	46.3	443	2	Q7TF26_9INFA	Q7tf26	influenza a	793	38	46.3	470	1	NRAM_IADBU	Q07570	influenza a
721	38	46.3	444	2	Q9I107_PSEAE	Q9i107	pseudomonas	794	38	46.3	470	1	NRAM_IADCH	Q07571	influenza a
722	38	46.3	445	2	Q7TF25_9INFA	Q7tf25	influenza a	795	38	46.3	470	1	NRAM_IADH2	Q07572	influenza a
723	38	46.3	445	2	Q2LFT4_9INFA	Q2lft4	influenza a	796	38	46.3	470	1	NRAM_IADU3	Q07599	influenza a
724	38	46.3	446	2	Q6GYV3_9INFA	Q6gyv3	influenza a	797	38	46.3	470	1	NRAM_IADU3	Q07576	influenza a
725	38	46.3	446	2	Q6GYV4_9INFA	Q6gyv4	influenza a	798	38	46.3	470	1	NRAM_IADU3	Q07578	influenza a
726	38	46.3	446	2	Q701T6_9INFA	Q701t6	influenza a	799	38	46.3	470	1	NRAM_IADU3	P31348	influenza a
727	38	46.3	446	2	Q701U1_9INFA	Q701u1	influenza a	800	38	46.3	470	1	NRAM_IADU3	P31349	influenza a
728	38	46.3	447	2	Q8HU2_9ARATH	Q8hu2	arabidopsis	801	38	46.3	470	2	Q595Z8_9INFA	Q595z8	influenza a
729	38	46.3	447	2	Q9FI18_9ARATH	Q9fi18	arabidopsis	802	38	46.3	470	2	Q5EDD2_9INFA	Q5edd2	influenza a
730	38	46.3	449	2	Q4FB58_9INFA	Q4fb58	influenza a	803	38	46.3	470	2	Q5WMA3_9INFA	Q5wma3	influenza a
731	38	46.3	449	2	Q5Q28_9INFA	Q5q28	influenza a	804	38	46.3	470	2	Q67218_9INFA	Q67218	influenza a
732	38	46.3	449	2	Q2L9W4_9INFA	Q2l9w4	influenza a	805	38	46.3	470	2	Q67219_9INFA	Q67219	influenza a
733	38	46.3	453	1	NRAM_IADU3	P03470	influenza a	806	38	46.3	470	2	Q67220_9INFA	Q67220	influenza a
734	38	46.3	454	1	NRAM_IADU3	P03468	influenza a	807	38	46.3	470	2	Q67221_9INFA	Q67221	influenza a
735	38	46.3	454	1	NRAM_IADU3	P03468	influenza a	808	38	46.3	470	2	Q67222_9INFA	Q67222	influenza a
736	38	46.3	454	2	Q6DX47_9INFA	Q6dx47	influenza a	809	38	46.3	470	2	Q7TG84_9INFA	Q7tg84	influenza a
737	38	46.3	454	2	Q8JU04_9INFA	Q8juu4	influenza a	810	38	46.3	470	2	Q7TG89_9INFA	Q7tg89	influenza a
738	38	46.3	455	2	Q94E25_9ARATH	Q94ez5	arabidopsis	811	38	46.3	470	2	Q7TG90_9INFA	Q7tg90	influenza a
739	38	46.3	455	2	Q212N6_9INFA	Q212n6	influenza a	812	38	46.3	470	2	Q7TG91_9INFA	Q7tg91	influenza a
740	38	46.3	460	2	Q8V3X8_9INFA	Q8v3x8	influenza a	813	38	46.3	470	2	Q8B199_9INFA	Q8b199	influenza a
741	38	46.3	462	1	FUCO_RAT	Q8V3X9	influenza a	814	38	46.3	470	2	Q8JMH6_9INFA	Q8jmh6	influenza a
742	38	46.3	462	2	Q2XK49_PSEPU	P17164	rattus norv	815	38	46.3	470	2	Q8JMH7_9INFA	Q8jmh7	influenza a
743	38	46.3	463	2	Q8V3X5_9INFA	Q8v3x5	influenza a	816	38	46.3	470	2	Q8JSD8_9INFA	Q8jstd8	influenza a
744	38	46.3	464	2	Q8M9B9_9POAL	Q8m9b9	mesanthemum	817	38	46.3	470	2	Q9WA95_9INFA	Q9wa95	influenza a
745	38	46.3	466	2	Q5W4S0_9ERRE	Q5w4s0	serenoa rep	818	38	46.3	470	2	Q21N42_9INFA	Q21n42	influenza a
746	38	46.3	466	2	Q2LFS6_9INFA	Q2lfs6	influenza a	819	38	46.3	470	2	Q21BHB_9INFA	Q21bhb	influenza a
747	38	46.3	468	2	Q5IR88_HUMAN	Q5ir88	homo sapien	820	38	46.3	475	2	Q4J622_SULAC	Q4j622	sulfolobus
748	38	46.3	468	2	Q6XV51_9INFA	Q6xv51	influenza a	821	38	46.3	475	2	Q8M9C1_9POAL	Q8m9c1	mesanthemum
749	38	46.3	469	2	Q4H2D1_9INFA	Q4h2d1	influenza a	822	38	46.3	477	2	Q8ID62_PLAF7	Q8id62	plasmodium
750	38	46.3	469	2	Q6DX95_9INFA	Q6dx95	influenza a	823	38	46.3	488	2	Q9MRA0_9CARY	Q9mra0	silene nuta
751	38	46.3	469	2	Q6GYU9_9INFA	Q6gyu9	influenza a	824	38	46.3	489	2	Q9MTY7_9ROSI	Q9mty7	salix retic
752	38	46.3	469	2	Q6GYV0_9INFA	Q6gyv0	influenza a	825	38	46.3	493	2	Q915R4_PSEAE	Q915r4	pseudomonas
753	38	46.3	469	2	Q6GYV5_9INFA	Q6gyv5	influenza a	826	38	46.3	496	2	Q9BA92_TRAFO	Q9ba92	trachycarpus
754	38	46.3	469	2	Q6GYV6_9INFA	Q6gyv6	influenza a	827	38	46.3	498	2	Q7HHY7_LICGR	Q7hh7	luciala gra
755	38	46.3	469	2	Q6GYV7_9INFA	Q6gyv7	influenza a	828	38	46.3	498	2	Q9B127_9LILI	Q9b127	pritchardio
756	38	46.3	469	2	Q6GYV8_9INFA	Q6gyv8	influenza a	829	38	46.3	498	2	Q9BA90_9LILI	Q9ba90	livistona s
757	38	46.3	469	2	Q6GYV9_9INFA	Q6gyv9	influenza a	830	38	46.3	498	2	Q9BA94_9LILI	Q9ba94	rhapis subt
758	38	46.3	469	2	Q6XV31_9INFA	Q6xv31	influenza a	831	38	46.3	498	2	Q9BA96_CHAHU	Q9ba96	chamaerops
759	38	46.3	469	2	Q6XV37_9INFA	Q6xv37	influenza a	832	38	46.3	502	2	Q445X3_SOLUS	Q445x3	solibacter
760	38	46.3	469	2	Q6XV54_9INFA	Q6xv54	influenza a	833	38	46.3	514	2	Q800L5_ANGJA	Q800l5	anguilla ja
761	38	46.3	469	2	Q6XV55_9INFA	Q6xv55	influenza a	834	38	46.3	516	2	Q342T2_RHOPA	Q342t2	rhodopseudo

835	38	46.3	516	2	Q447R5 SOLUS	Q447R5 solibacter	908	38	46.3	1032	2	Q82V80 PYRAE	Q82V80 pyrobaculum
836	38	46.3	520	2	Q2T769 BURTH	Q2T769 burkholderi	909	38	46.3	1032	2	Q7JLZ2 CAEBL	Q7JLZ2 caenorhabdi
837	38	46.3	520	2	P72647 SYNY3	P72647 synechocyst	910	38	46.3	1034	2	Q61J44 CAEBR	Q61J44 caenorhabdi
838	38	46.3	529	2	Q6ZJEB ORYSA	Q6ZJEB oryza sativ	911	38	46.3	1035	2	Q21079 CAEBL	Q21079 caenorhabdi
839	38	46.3	531	2	Q7WZ63 PACTO	Q7WZ63 nonomuraea	912	38	46.3	1040	1	Y043 CAEBL	Y043 caenorhabdi
840	38	46.3	538	2	Q443S5 SOLUS	Q443S5 solibacter	913	38	46.3	1053	2	Q3VV26 PROAE	Q3VV26 prosthecoch
841	38	46.3	541	2	Q340T4 RHOPA	Q340T4 rhodopsendo	914	38	46.3	1054	2	Q5RSN8 PONGY	Q5RSN8 pongo pygma
842	38	46.3	544	2	Q6N6G0 RHOPA	Q6N6G0 rhodopsendo	915	38	46.3	1073	2	Q69Z56 MOUSE	Q69Z56 mus musculu
843	38	46.3	554	2	Q89LD3 BRAJA	Q89LD3 bradyrhizob	916	38	46.3	1097	2	Q9HGT2 CANAL	Q9HGT2 candida alb
844	38	46.3	556	2	Q45233 BRAJA	Q45233 bradyrhizob	917	38	46.3	1114	2	Q4SXG7 TETNG	Q4SXG7 tetradodon n
845	38	46.3	560	2	Q4HXW1 GIBZE	Q4HXW1 gibberella	918	38	46.3	1117	2	Q5IRH9 HUMAN	Q5IRH9 homo sapien
846	38	46.3	560	2	Q73NKO TREDE	Q73NKO treponema d	919	38	46.3	1141	2	Q2WQJ7 HUMAN	Q2WQJ7 homo sapien
847	38	46.3	567	2	Q8KCI0 CHLTE	Q8KCI0 chlorobium	920	38	46.3	1187	2	Q5EMX8 PPOTV	Q5EMX8 plum pox vi
848	38	46.3	570	2	Q354Q9 BRAD	Q354Q9 bradyrhizob	921	38	46.3	1209	2	Q7XA27 SOLBU	Q7XA27 solanum bul
849	38	46.3	576	2	Q46Q20 RALEU	Q46Q20 ralstonia e	922	38	46.3	1226	2	Q4J5B1 AZOVI	Q4J5B1 azotobacter
850	38	46.3	578	2	Q60J20 CAEBR	Q60J20 caenorhabdi	923	38	46.3	1307	2	Q49AJ0 HUMAN	Q49AJ0 homo sapien
851	38	46.3	579	2	Q82P17 STRAW	Q82P17 streptomyc	924	38	46.3	1358	2	Q6FVM9 CANGA	Q6FVM9 candida gla
852	38	46.3	585	2	Q3FQQ3 9BURK	Q3FQQ3 rhodofera	925	38	46.3	1366	1	CO1A2 CANFA	Q46392 canis famila
853	38	46.3	591	2	Q9ZG10 THIFE	Q9ZG10 thioabacillu	926	38	46.3	1426	2	Q40868 LEIMA	Q40868 leishmania
854	38	46.3	597	2	Q3B6T8 ARATH	Q3B6T8 arabidopsis	927	38	46.3	1453	1	CO1A1 CHICK	P02457 gallus gall
855	38	46.3	597	2	Q5M7M1 XENTR	Q5M7M1 xenopus tro	928	38	46.3	1469	2	Q36VQ6 RHOPA	Q36VQ6 rhodopsendo
856	38	46.3	601	2	Q3QMD6 RHOB	Q3QMD6 silicibacte	929	38	46.3	1565	2	Q54CE3 DICDI	Q54CE3 dictyosteli
857	38	46.3	602	2	Q4P939 USTMA	Q4P939 ustilago ma	930	38	46.3	1737	2	Q37E12 RHOPA	Q37E12 rhodopsendo
858	38	46.3	613	2	Q96794 DROME	Q96794 drosophila	931	38	46.3	1739	2	Q3SSE5 NITWN	Q3SSE5 nitrobacter
859	38	46.3	613	2	Q9VJG0 DROME	Q9VJG0 drosophila	932	38	46.3	1867	2	Q4RI01 TETNG	Q4RI01 tetradodon n
860	38	46.3	621	2	Q64FAB MYAXA	Q64FAB myxococcus	933	38	46.3	3096	2	Q4GXV0 PPOTV	Q4GXV0 lily mottle
861	38	46.3	623	2	Q3H8C6 TRIER	Q3H8C6 trichodesmi	934	38	46.3	3140	1	POLG PPVRA	P17767 p genome po
862	38	46.3	625	2	Q826C5 STRAW	Q826C5 streptomyc	935	38	46.3	3140	1	POLG PPVSK	Q84934 p genome po
863	38	46.3	627	2	Q54G06 DICDI	Q54G06 dictyosteli	936	38	46.3	3140	2	Q2VQ18 PPOTV	Q2VQ18 plum pox vi
864	38	46.3	635	2	Q3PM99 NITHA	Q3PM99 nitrobacter	937	38	46.3	3140	2	Q2VQ19 PPOTV	Q2VQ19 plum pox vi
865	38	46.3	637	2	Q3EE13 ACTSC	Q3EE13 actinobacil	938	38	46.3	3140	2	Q2VQ20 PPOTV	Q2VQ20 plum pox vi
866	38	46.3	644	2	Q9ECM7 HUMAN	Q9ECM7 homo sapien	939	38	46.3	3140	2	Q2VQ21 PPOTV	Q2VQ21 plum pox vi
867	38	46.3	659	2	Q3DY01 CHLAU	Q3DY01 chloroflexu	940	38	46.3	3140	2	Q2VQ22 PPOTV	Q2VQ22 plum pox vi
868	38	46.3	661	2	Q2V3C6 ARATH	Q2V3C6 arabidopsis	941	38	46.3	3140	2	Q2VQ23 PPOTV	Q2VQ23 plum pox vi
869	38	46.3	662	1	KAT3 ARATH	P92960 arabidopsis	942	38	46.3	3140	2	Q2VQ24 PPOTV	Q2VQ24 plum pox vi
870	38	46.3	663	2	Q2ZWC5 BACLD	Q6ZWC5 bacillus li	943	38	46.3	3140	2	Q52V18 PPOTV	Q52V18 plum pox vi
871	38	46.3	665	2	Q5KX8 BACLD	Q65KX8 bacillus li	944	38	46.3	3140	2	Q52V19 PPOTV	Q52V19 plum pox vi
872	38	46.3	683	2	Q5YP51 NOCFA	Q5YP51 nocardia fa	945	38	46.3	3140	2	Q80S19 PPOTV	Q80S19 plum pox vi
873	38	46.3	688	2	Q4CBJ3 CLOTM	Q4CBJ3 clostridium	946	38	46.3	3140	2	Q80S20 PPOTV	Q80S20 plum pox vi
874	38	46.3	689	2	Q2RLK6 MOOTH	Q2RLK6 moorella th	947	38	46.3	3140	2	Q84925 PPOTV	Q84925 plum pox vi
875	38	46.3	696	2	Q6FVL3 CANGA	Q6FVL3 candida gla	948	38	46.3	3140	2	Q98WJ5 PPOTV	Q98WJ5 plum pox vi
876	38	46.3	705	2	Q6X4V5 GOSHI	Q6X4V5 gossypium h	949	38	46.3	3140	2	Q9PYF0 PPOTV	Q9PYF0 plum pox vi
877	38	46.3	744	2	Q4RLD6 TETNG	Q4RLD6 tetradodon n	950	38	46.3	3141	2	Q52V21 PPOTV	Q52V21 plum pox vi
878	38	46.3	747	1	PPN1 YARLI	Q6CEE7 yarrowia li	951	38	46.3	3587	2	Q6RKK6 GIBMO	Q6RKK6 gibberella
879	38	46.3	750	2	Q9N1N8 HUMAN	Q9N1N8 homo sapien	952	38	46.3	4684	2	Q7R1S4 GIALA	Q7R1S4 giardia lam
880	38	46.3	753	2	Q5YU9 NOCFA	Q5YU9 nocardia fa	953	38	46.3	6939	2	Q2T1K7 BURTH	Q2T1K7 burkholderi
881	38	46.3	760	2	Q4AIR7 9CHLB	Q4AIR7 chlorobium	954	37.5	45.7	108	2	Q2T1K7 BURTH	Q2T1K7 burkholderi
882	38	46.3	763	2	Q4IKV5 METBU	Q4IKV5 methanococc	955	37.5	45.7	236	2	Q3B0E5 SYN9S	Q3B0E5 synechococc
883	38	46.3	799	2	Q82KF5 STRAW	Q82KF5 streptomyc	956	37.5	45.7	244	2	Q4HSU6 PDEIO	Q4HSU6 deinocococc
884	38	46.3	810	1	NELL1 RAT	Q62919 rattus norv	957	37.5	45.7	256	2	Q7U9S0 SYNXP	Q7U9S0 synechococ
885	38	46.3	810	2	Q2VWQ2 MOUSE	Q2VWQ2 mus musculu	958	37.5	45.7	279	2	Q8U7B4 AGRT5	Q8U7B4 agrobacteri
886	38	46.3	822	2	Q5D1Z2 PPOTV	Q5D1Z2 plum pox vi	959	37.5	45.7	282	2	Q7NU47 CHRVO	Q7NU47 chromobacte
887	38	46.3	824	2	Q7VLM9 HAEDU	Q7VLM9 haemophilus	960	37.5	45.7	290	2	Q9P827 CANAL	Q9P827 candida alb
888	38	46.3	858	2	Q6FN07 CANGA	Q6FN07 candida gla	961	37.5	45.7	297	2	Q59RM2 CANAL	Q59RM2 candida alb
889	38	46.3	860	1	ATS6 HUMAN	Q9UKP5 homo sapien	962	37.5	45.7	297	2	Q59S23 CANAL	Q59S23 candida alb
890	38	46.3	860	2	Q6D8U4 ERWCT	Q6D8U4 erwinia car	963	37.5	45.7	304	2	Q7CV72 AGRT5	Q7CV72 agrobacteri
891	38	46.3	866	2	Q4U4W7 THEAU	Q4U4W7 thermoaacus	964	37.5	45.7	338	2	Q353P1 9GAMM	Q353P1 alkalilimni
892	38	46.3	907	2	Q988S0 CARAU	Q988S0 carassius a	965	37.5	45.7	345	2	Q7ME20 VIBVU	Q7ME20 vibrio vuln
893	38	46.3	913	2	Q2XE12 PSEPU	Q2XE12 pseudomonas	966	37.5	45.7	345	2	Q8D7X3 VIBVU	Q8D7X3 vibrio vuln
894	38	46.3	916	2	Q5BA51 EMENI	Q5BA51 aspergillus	967	37.5	45.7	360	2	Q2RF44 RHODR	Q2RF44 rhodospiril
895	38	46.3	929	2	Q7F8S7 ORYSA	Q7F8S7 oryza sativ	968	37.5	45.7	382	2	Q48BQ8 PSE14	Q48BQ8 pseudomonas
896	38	46.3	947	2	Q8TF46 HUMAN	Q8TF46 homo sapien	969	37.5	45.7	382	2	Q87U42 PSESM	Q87U42 pseudomonas
897	38	46.3	955	2	Q3B6W2 PELLD	Q3B6W2 pelodictyon	970	37.5	45.7	436	2	Q8BWH5 MOUSE	Q8BWH5 mus musculu
898	38	46.3	970	2	Q8C0S1 MOUSE	Q8C0S1 mus musculu	971	37.5	45.7	440	2	Q7V3Y1 PROMM	Q7V3Y1 prochloroco
899	38	46.3	971	2	Q8WTU9 HUMAN	Q8WTU9 homo sapien	972	37.5	45.7	507	2	Q4W907 ASPFU	Q4W907 aspergillus
900	38	46.3	971	2	Q5U2P0 RAT	Q5U2P0 rattus norv	973	37.5	45.7	574	2	Q84FK7 ENTAG	Q84FK7 enterobacte
901	38	46.3	976	2	Q73SX6 MYCPA	Q73SX6 mycobacteri	974	37.5	45.7	582	2	Q4DV56 TRYCR	Q4DV56 trypanosoma
902	38	46.3	983	2	Q5ZL00 CHICK	Q5ZL00 gallus gall	975	37.5	45.7	679	2	Q2U8S2 ASPOR	Q2U8S2 aspergillus
903	38	46.3	994	2	Q2JUC9 PACTO	Q2JUC9 frankia sp.	976	37.5	45.7	734	2	Q5CBP9 THENE	Q5CBP9 thermotoga
904	38	46.3	999	2	Q3C207 9VIRU	Q3C207 heterocapsa	977	37.5	45.7	735	2	Q5CBG4 9THEM	Q5CBG4 thermotoga
905	38	46.3	1003	2	Q39S98 GEOMG	Q39S98 geobacter m	978	37.5	45.7	747	2	Q4WIU2 ASPFU	Q4WIU2 aspergillus
906	38	46.3	1004	2	Q3C205 9VIRU	Q3C205 heterocapsa	979	37.5	45.7	772	2	Q89PP4 BRAJA	Q89PP4 bradyrhizob
907	38	46.3	1010	1	CLPP CHLEU	P42379 chlamydomon	980	37.5	45.7	1004	1	YD83 SCHPO	Y10408 schizosacch

981 37.5 45.7 1401 2 Q80YP9 mus musculus
 982 37.5 45.7 2581 2 Q8ZS63 anabaena sp
 983 37.5 45.7 2766 2 Q3VZE3 frankia sp.
 984 37 45.1 64 2 Q96GX8 homo sapien
 985 37 45.1 100 2 Q4WHB8 aspergillus
 986 37 45.1 101 2 Q3HJ08 trichodemi
 987 37 45.1 101 2 Q4C709 crocospaer
 988 37 45.1 106 2 Q4KL44 mouse
 989 37 45.1 108 2 Q3FHN0 burkholderi
 990 37 45.1 118 2 Q62WM3 bacillus li
 991 37 45.1 119 2 Q6Z355 oryza sativ
 992 37 45.1 119 2 Q7UJ28 rhodospirall
 993 37 45.1 121 2 Q8G3F9 bifidobacte
 994 37 45.1 129 2 Q6D326 erwinia car
 995 37 45.1 131 2 Q651A5 oryza sativ
 996 37 45.1 131 2 Q6J9R1 arabidopsis
 997 37 45.1 131 2 Q9LSX0 arath
 998 37 45.1 131 2 Q43U69 solus
 999 37 45.1 132 2 Q87Sf6 vibrio para
 1000 37 45.1 136 2 Q4B963 burkholderi

ALIGNMENTS

RESULT 1

HPSE HUMAN
 ID HPSE_HUMAN STANDARD; PRT; 543 AA.
 AC Q9Y251; Q53GE5; Q9UL39;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-NOV-1999, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Heparanase precursor [EC 3.2.-.-] (Heparanase-1) (Hpal) (Endo-
 DE glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
 DE subunit).
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRI, HPSE1, HSE1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=9935379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
 RA Seddon A.P., Giorgio N.A., Bohlen P.;
 RT "Cloning and functional expression of a human heparanase gene";
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 [2]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.
 RC TISSUE=Embryonic fibroblast;
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
 RA Toyoshima M., Nakajima M.;
 RT "Human heparanase. Purification, characterization, cloning, and
 RT expression.";
 RL J. Biol. Chem. 274:24153-24160(1999).
 [3]
 RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
 RP PubMed=10395325; DOI=10.1038/10518;
 RX Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
 RA Spector L., Pecker I.;
 RT "Mammalian heparanase: gene cloning, expression and function in tumor
 RT progression and metastasis";
 RL Nat. Med. 5:793-802(1999).
 [4]
 RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 [5]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage";
 RL Glycobiology 10:467-475(2000).
 [6]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
 [7]
 RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhanan M., Felix R.,
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Parekh R., Stubbsfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpal) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435(2003).
 [8]
 RN NUCLEOTIDE SEQUENCE [MRNA].
 RP Pinnal M.A., Semedo P.;
 RT "Cloned heparanase from MCF-7 cells.";
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 [9]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP TISSUE=Small intestine;
 RC Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
 RA Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 [10]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [11]
 RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RP PubMed=11123890; DOI=10.1021/bi002080p;
 RX Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Gready J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic
 RL endoglycosidase heparanase.";
 RN Biochemistry 39:15659-15667(2000).
 RP [12]
 RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
 RP ASN-217; ASN-238 AND ASN-459.
 RX PubMed:14573609; DOI=10.1074/jbc.M300541200;
 RA Sinizu S., Ishida K., Wierzb M.K., Osada H.;
 RT Secretion of heparanase protein is regulated by glycosylation in
 RL human tumor cell lines.";
 RN J. Biol. Chem. 279:2697-2703(2004).
 RP [13]
 RP SUBCELLULAR LOCATION.
 RX PubMed:15292202; DOI=10.1074/jbc.M402131200;
 RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
 RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
 RT "Heparanase uptake is mediated by cell membrane heparan sulfate
 RL proteoglycans.";
 RN J. Biol. Chem. 279:44084-44092(2004).
 RP [14]
 RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
 RX PubMed:15659389; DOI=10.1074/jbc.M413370200;
 RA Aboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
 RA Elgavish S., Peretz T., Vlodavsky I.;
 RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
 RL human proheparanase.";
 RN J. Biol. Chem. 280:13568-13575(2005).
 RP [16]
 RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
 RX PubMed:15760902; DOI=10.1074/jbc.M414546200;
 RA Levy-Adam F., Aboud-Jarrous G., Guerrini M., Beccati D.,
 RA Vlodavsky I., Ilan N.;
 RT "Identification and characterization of heparin/heparan sulfate
 RL binding domains of the endoglycosidase heparanase.";
 RN J. Biol. Chem. 280:20457-20466(2005).
 RP [17]
 RP VARIANT SER-260.
 RX PubMed:15334672;
 RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
 RA Shi L.H., Shen H.W., Xu B.;
 RT "Heparanase mRNA expression and point mutation in hepatocellular
 RL carcinoma.";
 RN World J. Gastroenterol. 10:2795-2799(2004).
 RP [18]
 RP FUNCTION: Endoglycosidase which is a cell surface and
 RL extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Due to its contribution to metastasis and
 CC angiogenesis, it is considered to be a potential target for anti-
 CC cancer therapies.
 CC [19]
 CC ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
 CC lower extent, by heparin and sulfamin and activated by calcium and
 CC magnesium (By similarity).
 CC [20]
 CC BIOPHYSICOCHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 4-6;
 CC [21]
 CC SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products.
 CC [22]
 CC SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium.
 CC [23]
 CC TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
 CC weakly expressed in lymph node, thymus, peripheral blood
 CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.
 CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme.
 CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility.
 Query Match 100.0%; Score 82; DB 1; Length 543;
 Best Local Similarity 100.0%; Pred. No. 3.8e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RPQKVKWLGETSSAY 15
 DB 334 RPQKVKWLGETSSAY 348
 |||||
 RESULT 2
 Q333X5 SPAJD PRELIMINARY; PRT; 558 AA.
 AC Q333X5;
 DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 3.
 DE Heparanase.
 GN Name:hpa;
 OS Spalax judaei (Blind subterranean mole rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Spalacidae; Spalacinae; Spalax.
 OC NCBI_TaxId=134510;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE-Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 RL cloning and identification of a novel splice variant.";
 RN Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
 RP [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE-Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 RL cloning and identification of a novel splice variant.";
 RN Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AM085494; CAJ30021.1; -; mRNA.
 CC SEQUENCE 558 AA; 62737 MW; 07BAF8F55849EEE7 CRC64;
 Query Match 100.0%; Score 82; DB 2; Length 558;
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RPQKVKWLGETSSAY 15
 DB 349 RPQKVKWLGETSSAY 363
 |||||
 RESULT 3
 Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
 AC Q333X6;
 DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 3.
 DE Heparanase.
 GN Name:hpa;
 OS Spalax judaei (Blind subterranean mole rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2]
RP NUCLEOTIDE SEQUENCE.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match 100.0%; Score 82; DB 2; Length 574;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGTTSSAY 15
Db |||||
365 RPKKKVWLGTTSSAY 379

RESULT 4
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [1]
RP NUCLEOTIDE SEQUENCE.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCBADD99DE CRC64;

Query Match 100.0%; Score 82; DB 2; Length 574;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGTTSSAY 15
Db |||||
365 RPKKKVWLGTTSSAY 379

RESULT 5
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.

```

```

AC Q333X8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48BEFEC7D0BCB34 CRC64;

Query Match 100.0%; Score 82; DB 2; Length 574;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGTTSSAY 15
Db |||||
365 RPKKKVWLGTTSSAY 379

RESULT 6
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

```


DR EMBL; AM085490; CAJ30017.1; -; mRNA.
 SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;
 Query March 100.0%; Score 82; DB 2; Length 574;
 Best Local Similarity 100.0%; Pred. No. 4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RPQKKVWLGESSAY 15
 |||||
 DB 365 RPQKKVWLGESSAY 379
 RESULT 7
 HPSE MOUSE
 ID HPSE MOUSE STANDARD; PRT; 535 AA.
 AC Q6VGGZ1; Q8K3K3;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 11-OCT-2005, sequence version 2.
 DT 07-MAR-2006, entry version 13.
 DE Heparanase precursor (EC 3.2.2.-) (Endo-glucuronidase) [Contains:
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
 GN Name=Hps; Synonyms=Hpa;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridea; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RC STRAIN=SJL/J; TISSUE=Spleen;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
 RA Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis";
 RL Nat. Med. 5:803-809(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
 RP GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND
 RP SUBUNIT.
 RC STRAIN=FVB; TISSUE=Embryo;
 RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
 RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
 RA Plata A., Zhou Q., Ludwig D., Bohnen P., Kussie P.;
 RT "Cloning, expression, and purification of mouse heparanase";
 RL Protein Expr. Purif. 26:425-431(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
 RA Gong F., Jemth P., Galvis M.L.E., Vlodaysky I., Horner A., Lindahl U.,
 RA Li J.-P.;
 RT "Processing of macromolecular heparin by heparanase";
 RL J. Biol. Chem. 278:35152-35158(2003).
 RN [4]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
 RX PubMed=16141072; DOI=10.1126/science.112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Humanecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,

RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 RA Matsuoka H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
 RA Petkovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sees L., Sheng Y.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusci V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
 RA Nishio T., Okada M., Plesky C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome";
 RL Science 309:1559-1563(2005).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Contributes to metastasis and angiogenesis
 CC (By similarity).
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and
 CC magnesium (By similarity). Inhibited by laminarin sulfate and, to
 CC a lower extent, by heparin and sulfamin.
 CC -!- BIOPHYSICO-CHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 5;
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products.
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium
 CC (By similarity).
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL; AF359507; AAQ15188.1; -; mRNA.
 CC EMBL; AY077467; AAL76083.1; -; mRNA.
 CC EMBL; AY151051; AAN41636.1; -; mRNA.
 CC EMBL; AK040471; BAC30600.1; -; mRNA.
 CC EMBL; AK154628; BAE32725.1; -; mRNA.
 CC Ensembl; ENSMUSG0000035273; Mus musculus.
 CC MGI; MGI:1343124; Hpsa.
 CC GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
 CC InterPro; IPR005199; Glyco_hydro_79_N.
 CC Pfam; PF03662; Glyco_hydro_79n; 1.
 CC Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
 KW Magnesium; Membrane; Signal.
 FT SIGNAL 1 27 By similarity.
 FT CHAIN 28 101 Heparanase 8 kDa subunit.
 FT /FTID=PRO_0000042263.
 FT PROPEP 102 149 Linker peptide (By similarity).

```

FT CHAIN 150 535 /FTid=PRO_0000042264.
FT Heparanase 50 kDa subunit.
FT /FTid=PRO_0000042265.
FT Heparin/HS-binding (By similarity).
FT REGION 150 154 Heparin/HS-binding (By similarity).
FT ACT_SITE 262 272 Proton donor (Potential).
FT ACT_SITE 217 217 Nucleophile (Potential).
FT CARBOHYD 335 335 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 154 154 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 192 192 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 209 209 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 230 230 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 451 451 K -> R (in Ref. 3).
FT CONFLICT 206 206 W -> S (in Ref. 3).
FT CONFLICT 212 212 NGS -> DGL (in Ref. 1, 2 and 4).
FT CONFLICT 230 232 E -> K (in Ref. 3).
FT CONFLICT 335 335 G -> A (in Ref. 3).
FT CONFLICT 342 342 Y -> H (in Ref. 1, 2 and 4).
FT CONFLICT 455 455 V -> I (in Ref. 1, 2 and 4).
FT CONFLICT 531 531
FT SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;

Query Match 93.9%; Score 77; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
Db 327 PGKKVWLGETSSAY 340

RESULT 8
HPSE RAT
ID HPSE RAT STANDARD; PRT; 536 AA.
AC Q9IPL; Q9QZF8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Heparanase precursor [EC 3.2.-.-] (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
DE Name=Hpse; Synonyms=Hep;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
OX [1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RP MEDLINE=22194309; PubMed=10395326; DOI=10.1038/10525;
RX Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line.";
RL J. Biol. Chem. 277:32459-32465 (2002).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes

```

```

CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF359508; AAQ15189.1; -; mRNA.
DR EMBL; AF184967; AAF04563.1; -; mRNA.
DR RGD; 61969; Hpse.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 Linker peptide (By similarity).
FT CHAIN 151 536 Heparanase 50 kDa subunit.
FT REGION 151 155 Heparin/HS-binding (By similarity).
FT ACT_SITE 263 273 Heparin/HS-binding (By similarity).
FT ACT_SITE 218 218 Proton donor (Potential).
FT CARBOHYD 336 336 Nucleophile (Potential).
FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 93.9%; Score 77; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
Db 328 PGKKVWLGETSSAY 341

RESULT 9
HPSE CHICK
ID HPSE CHICK STANDARD; PRT; 523 AA.
AC Q90TK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
DE Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;

```

RT	"Expression pattern and secretion of human and chicken heparanase are determined by their signal peptide sequence.";										
RL	J. Biol. Chem. 276:29178-29187(2001).										
CC	-!- FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans (By similarity).										
CC	-!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted (By similarity).										
CC	-!- PTM: N-glycosylated (By similarity).										
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.										
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms										
CC	Distributed under the Creative Commons Attribution-NoDerivs License										
CC	-----										
DR	ENBL; AY037007; AAK82648.1; -; mRNA.										
DR	Ensembl; ENSGALG00000011203; Gallus gallus.										
DR	InterPro: IPR005199; Glyco_hydro_79_N.										
DR	Pfam: PF03662; Glyco_hydro_79n; 1.										
KW	Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.										
FT	SIGNAL	1	18								
FT	CHAIN	19	523								
FT	REGION	137	141								
FT	REGION	250	260								
FT	ACT_SITE	204	204								
FT	ACT_SITE	323	323								
FT	CARBOHYD	141	141								
FT	CARBOHYD	196	196								
FT	CARBOHYD	436	436								
FT	CARBOHYD	439	439								
SQ	SEQUENCE	523 AA;	58386 MW;	8EB0B7B18C9BF881	CRC64;						
Query Match	89.0%;	Score 73;	DB 1;	Length 523;							
Best Local Similarity	92.9%;	Pred. No. 0.0013;									
Matches 13;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps							
Qy	2	PGKKVWLGETSSAY 15									
Db	315	PGKKVWLGETGSAY 328									
RESULT 10											
Q4SYF6	TETNG										
ID	Q4SYF6	TETNG	PRELIMINARY;	PRT;	533	AA.					
AC	Q4SYF6;										
DT	19-JUL-2005,	integrated into UniProtKB/TrEMBL.									
DT	19-JUL-2005,	sequence version 1.									
DT	07-FEB-2006,	entry version 4.									
DE	Chromosome undetermined SCAF12073, whole genome shotgun sequence.										
DE	(Fragment).										
GN	ORFNames=GSTENG00010356001;										
OS	Tetraodon nigroviridis (Green puffer).										
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;										
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;										
OC	Tetraodontoidea; Tetraodontidae; Tetraodon.										
OX	NCBI_TaxID=99893;										
RN	[1]										
RP	NUCLEOTIDE SEQUENCE.										
RX	PubMed=15496914; DOI=10.1038/nature03025;										
RA	Jailton O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,										
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,										
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,										
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,										
RA	Anthouard V., Jubin C., Castelli V., Katinka M., Vachexie B.,										
RA	Blemond C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,										
RA	Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,										
RA	Parra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,										
RA	Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,										
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,										
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,										
RA	Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.										

"Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype."; [2]
RN Nature 431:946-957(2004).
RN NCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL; CAAE01012073; CAF94326.1; -; Genomic_DNA.
CC FT NON TER 1 533
CC FT NON TER 533 533
CC SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;

Query Match 86.6%; Score 71; DB 2; Length 533;
Best Local Similarity 92.9%; Pred. No. 0.0029;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PGKXWLGTSY 15
||| |||||
Db 291 PGKFWLGTSY 304

RESULT 11
HPSE_BOVIN
ID HPSE_BOVIN STANDARD; PRT; 545 AA.
AC QNMYO;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor [EC 3.2.1.-] [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
GN Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
OX 1
[1]
RN NCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;
RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -!- PMW: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant

CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).

CC -(- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility (By similarity).

CC -(- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL: AF281160; AAF87301.2; -; mRNA.
 DR InterPro: IPR005199; Glyco_hydro_79_N.
 DR Pfam: PF03662; Glyco_hydro_79n; I.
 DR Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
 KW Signal.

FT SIGNAL 1 37 By similarity.
 FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
 FT FT /FTid=PRO_0000042256.
 FT PROPEP 112 159 Linker peptide.
 FT FT /FTid=PRO_0000042257.
 FT CHAIN 160 545 Heparanase 50 kDa subunit (By
 FT similarity).
 FT FT /FTid=PRO_0000042258.

FT REGION 160 164 Heparin/HS-binding (Potential).
 FT REGION 272 282 Heparin/HS-binding (Potential).
 FT ACT_SITE 227 227 Proton donor (Potential).
 FT ACT_SITE 345 345 Nucleophile (Potential).
 FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD85B933 CRC64;

Query Match 82.9%; Score 68; DB 1; Length 545;
 Best Local Similarity 86.7%; Pred. No. 0.0099;
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPKKVKWLGETSSAY 15
 DB 336 RPLKVKWLGETSSAF 350
 || |||||

RESULT 12

Q4TGC8_TETNG NUCLEOTIDE SEQUENCE.
 ID Q4TGC8_TETNG PRELIMINARY; PRT; 255 AA.
 AC Q4TGC8_TETNG, integrated into UniProtKB/TrEMBL.
 DT 19-JUL-2005, sequence version 1.
 DT 07-FEB-2006, entry version 4.
 DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
 DE (Fragment).
 DE ORFNames=GSTENG0001168001;
 OS Tetraodon nigroviridis (Green puffer).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontoidea; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN NUCLEOTIDE SEQUENCE.
 RP PubMed=15496914; DOI=10.1038/nature03025;
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Guzy J.,
 RA Parra G., Lardier G., Chappell C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RN Nature 431:946-957(2004).
 RL NUCLEOTIDE SEQUENCE.
 RP Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -(- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RL NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -(- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL: CAAB01003783; CAF88054.1; -; Genomic_DNA.
 FT NON_TER 1 255
 FT SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;
 SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

Query Match 78.0%; Score 64; DB 2; Length 255;
 Best Local Similarity 92.3%; Pred. No. 0.022;
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 RPKKVKWLGETSSA 14
 DB 171 RPKKVKWLGETSSA 183
 |||||

RESULT 13

Q4T880_TETNG NUCLEOTIDE SEQUENCE.
 ID Q4T880_TETNG PRELIMINARY; PRT; 597 AA.
 AC Q4T880_TETNG, integrated into UniProtKB/TrEMBL.
 DT 19-JUL-2005, sequence version 1.
 DT 07-FEB-2006, entry version 4.
 DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
 DE ORFNames=GSTENG0003868001;
 OS Tetraodon nigroviridis (Green puffer).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontoidea; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN NUCLEOTIDE SEQUENCE.
 RP PubMed=15496914; DOI=10.1038/nature03025;
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Guzy J.,
 RA Parra G., Lardier G., Chappell C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RN Nature 431:946-957(2004).
 RL NUCLEOTIDE SEQUENCE.
 RP Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -(- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL: CAAB01007180; CAF89852.1; -; Genomic_DNA.
 FT NON_TER 597 597

```
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
Query Match 61.0%; Score 50; DB 2; Length 597;
Best Local Similarity 64.3%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
|||||||
|:
Db 375 PGKKVWLGGLGPAW 388

RESULT 14
Q2IPR2_9DELT PRELIMINARY; PRT; 586 AA.
AC Q2IPR2_
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Thiamine pyrophosphate enzyme.
GN ORFNames=Adelh 1015;
OS Anaeromyxobacter dehalogenans 2CP-C.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cytoacterineae; Myxococcaceae; Anaeromyxobacter.
OX NCBI_TaxID=290397;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Brettin T., Bruce D., Han C.,
RA Tapia R., Gilna P., Kiss H., Schmutz J., Larimer F., Land M.,
RA Kyrpides N., Anderson J., Sanford R.A., Ritalahti K.M., Thomas H.S.,
RA Kirby J.R., Zhulin I.B., Loeffler F.E., Richardson P.;
RT "Complete sequence of Anaeromyxobacter dehalogenans 2CP-C."
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; CP000251; ABC80790.1; -; Genomic DNA.
SQ SEQUENCE 586 AA; 62020 MW; 39D77C7222A18EF7 CRC64;

Query Match 59.8%; Score 49; DB 2; Length 586;
Best Local Similarity 58.8%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 1 RPKKVWL--GETSSAY 15
|||||||
|:
Db 466 RPGEVWLLYGDGSCAY 482

RESULT 15
Q36KN3_MARHY PRELIMINARY; PRT; 197 AA.
AC Q36KN3;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Phosphonate metabolism.
GN ORFNames=MaquidRAFT_4132;
OS Marinobacter aquaeolei VT8.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Marinobacter.
OX NCBI_TaxID=351348;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Marinobacter aquaeolei VT8."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.

SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
Query Match 61.0%; Score 50; DB 2; Length 597;
Best Local Similarity 64.3%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
|||||||
|:
Db 375 PGKKVWLGGLGPAW 388

RESULT 14
Q2IPR2_9DELT PRELIMINARY; PRT; 586 AA.
AC Q2IPR2_
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Thiamine pyrophosphate enzyme.
GN ORFNames=Adelh 1015;
OS Anaeromyxobacter dehalogenans 2CP-C.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cytoacterineae; Myxococcaceae; Anaeromyxobacter.
OX NCBI_TaxID=290397;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Brettin T., Bruce D., Han C.,
RA Tapia R., Gilna P., Kiss H., Schmutz J., Larimer F., Land M.,
RA Kyrpides N., Anderson J., Sanford R.A., Ritalahti K.M., Thomas H.S.,
RA Kirby J.R., Zhulin I.B., Loeffler F.E., Richardson P.;
RT "Complete sequence of Anaeromyxobacter dehalogenans 2CP-C."
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; CP000251; ABC80790.1; -; Genomic DNA.
SQ SEQUENCE 586 AA; 62020 MW; 39D77C7222A18EF7 CRC64;

Query Match 59.8%; Score 49; DB 2; Length 586;
Best Local Similarity 58.8%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 1 RPKKVWL--GETSSAY 15
|||||||
|:
Db 466 RPGEVWLLYGDGSCAY 482

RESULT 15
Q36KN3_MARHY PRELIMINARY; PRT; 197 AA.
AC Q36KN3;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Phosphonate metabolism.
GN ORFNames=MaquidRAFT_4132;
OS Marinobacter aquaeolei VT8.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Marinobacter.
OX NCBI_TaxID=351348;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Marinobacter aquaeolei VT8."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
```

```
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Marinobacter aquaeolei VT8."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AALG01000036; EAO97515.1; -; Genomic DNA.
DR GO; GO:0015716; P:phosphonate transport; IEA.
SQ SEQUENCE 197 AA; 21261 MW; E86FF9188614ACC3 CRC64;

Query Match 58.5%; Score 48; DB 2; Length 197;
Best Local Similarity 61.5%; Pred. No. 9.7;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSA 14
|||||
|:
Db 42 PGESVWLADTDSA 54

RESULT 16
Q3QDK4_9GAMM PRELIMINARY; PRT; 616 AA.
ID Q3QDK4_9GAMM
AC Q3QDK4;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Ferredoxin:Oxidoreductase FAD/NAD(P)-binding.MOSC:MOSC, N-terminal
DE beta barrel:Oxidoreductase FAD-binding region.
GN ORFNames=SamadRAFT_1894;
OS Shewanella amazonensis SB2B.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=326297;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB2B;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella amazonensis SB2B."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AAIN01000034; EAN37953.1; -; Genomic DNA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006058; 2Fe2S_fd_BS.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR005302; MOSC.
```

```
DR InterPro; IPR005303; MOSC.N.
DR InterPro; IPR008333; Oxred FAD bd.
DR InterPro; IPR001433; Oxred FAD NAD bd.
DR InterPro; IPR001221; Phe hydroxylase.
DR Pfam; PF00970; FAD binding_6; 1.
DR Pfam; PF00111; Fer2; 1.
DR Pfam; PF03473; MOSC; 1.
DR Pfam; PF03476; MOSC.N; 1.
DR Pfam; PF00175; NAD Binding 1; 1.
DR PRINTS; PR00410; PHEHYDRXLASE.
DR PROSITE; PS00197; 2FE2S_FER_1; 1.
DR PROSITE; PS1085; 2FE2S_FER_2; 1.
SQ SEQUENCE 616 AA; 67299 MW; EDB349A10D7494A4 CRC64;

Query Match 58.5%; Score 48; DB 2; Length 616;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
| : ||||| :
Db 122 PARLIWLGETSNRF 135

RESULT 17
Q41J33 METBU
ID Q41J33 METBU PRELIMINARY; PRT; 334 AA.
AC Q41J33;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Trimethylamine methyltransferase.
GN OFRNAmes-MburaDRAFT_0457;
OS Methanococcoides burtonii DSM 6242.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanococcoides.
OX NCBI_TaxID=259564;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 6242;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Methanococcoides
RT burtonii DSM 6242.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=DSM 6242;
RC US DOE Joint Genome Institute (JGI-ORNL);
RG Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Methanococcoides burtonii
RT DSM 6242.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[3]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=DSM 6242;
RC US DOE Joint Genome Institute (JGI-PGF);
RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA Submittted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AADH02000014; EAM99398.1; -; Genomic DNA.
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0015948; P:methanogenesis; IEA.
DR InterPro; IPR010426; MTTB.
DR Pfam; PF06253; MTTB; 1.
```

```
KW Methyltransferase; Transferase.
SQ SEQUENCE 334 AA; 36229 MW; 5C8FCB81F0FE0D2A CRC64;

Query Match 57.3%; Score 47; DB 2; Length 334;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
| : ||||| :
Db 284 PGAKVWYGSSTTAF 297

RESULT 18
Q54359 STRLN
ID Q54359 STRLN PRELIMINARY; PRT; 426 AA.
AC Q54359;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 21-FEB-2006, entry version 25.
DE LmbF protein.
GN Name=lmbF;
OS Streptomyces lincolnensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1915;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=78-11;
RX MEDLINE=96026646; PubMed=8577249;
RA Peschke U., Schmidt H., Zhang H.Z., Piepersberg W.;
RT "Molecular characterization of the lincomycin-production gene cluster
RT of Streptomyces lincolnensis 78-11.";
RL Mol. Microbiol. 16:1137-1156(1995).
CC -1- COPACTOR: Pyridoxal phosphate (By similarity).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; X79146; CAA55752.1; -; Genomic DNA.
DR PIR; S69815; S44953.
DR GO; GO:0016769; F:transferase activity, transferring nitrogen. .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR004839; Aminotrans I/II.
DR Pfam; PF00155; Aminotran 1.2; 2.
DR Pyridoxal phosphate; Transferase.
KW SEQUENCE 426 AA; 46681 MW; 553250E8E083BA2 CRC64;

Query Match 57.3%; Score 47; DB 2; Length 426;
Best Local Similarity 60.0%; Pred. No. 33;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RPKKVWLGETSSAY 15
| : ||||| :
Db 395 RPWFKVWLGRDSSVF 409

RESULT 19
Q6LAL3_9POTV
ID Q6LAL3_9POTV PRELIMINARY; PRT; 3143 AA.
AC Q6LAL3;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Polyprotein.
OS Plum pox virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=12211;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Sweet Cherry;
RX MEDLINE=98163291; PubMed=9504763; DOI=10.1016/S0166-0934(97)00158-4;
RA Crescenzi A., d'Aquino L., Nuzzaci M., Ostuni A., Bavoso A., Comes S.,
```

RA De Stradis A., Piazzolla P.;
RT "Production of strain specific antibodies against a synthetic
RT polypeptide corresponding to the N-terminal region of the plum pox
RL polyvirus coat protein."; J. Virol. Methods 69:181-189(1997).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Sweet Cherry;
RX MEDLINE=22939292; PubMed=14579174; DOI=10.1007/s00705-003-0175-9;
RA Fanigliulo A., Comes S., Maiss E., Piazzolla P., Crescenzi A.;
RT "The complete nucleotide sequence of Plum pox virus isolates from
RT sweet (PPV-SwC) and sour (PPV-Soc) cherry and their taxonomic
RL relationships within the species."; Arch. Virol. 148:2137-2153(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; Y09851; CAF02291.1; -; Genomic_RNA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002540; Pept_S30_Poty_P1.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001456; Peptidase_C6.
DR InterPro; IPR001592; Poty_coat.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00851; Peptidase_C6; 1.
DR Pfam; PF01577; Peptidase_S30; 1.
DR Pfam; PF00767; Poty_coat; 1.
DR Pfam; PF00680; RdRP_1; 1.
DR PRINTS; PR00966; NIAPOTYPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR CAPSID protein; Polyprotein. P1 protein.
FT CHAIN 1 308 HC protein.
FT CHAIN 309 766 P3 protein.
FT CHAIN 767 1116 6K1 protein.
FT CHAIN 1117 1168 6K2 protein.
FT CHAIN 1169 1803 N1a protein.
FT CHAIN 1804 1856 N1a protein.
FT CHAIN 1857 2293 replicase.
FT CHAIN 2294 2811 coat protein.
FT CHAIN 2812 3143
SQ SEQUENCE 3143 AA; 458251AA2279108E CRC64;

Query Match 57.3%; Score 47; DB 2; Length 3143;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PGKKVWLGSTSS 13
| | | | | : : : |
Db 311 PGKKFWVGFTNS 322

RESULT 20
Q6Y3X4_9POTV PRELIMINARY; PRT; 3143 AA.
ID Q6Y3X4_9POTV

AC Q6Y3X4;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Polyprotein.
OS Plum pox virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=12211;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22939292; PubMed=14579174; DOI=10.1007/s00705-003-0175-9;
RA Fanigliulo A., Comes S., Maiss E., Piazzolla P., Crescenzi A.;
RT "The complete nucleotide sequence of Plum pox virus isolates from
RT sweet (PPV-SwC) and sour (PPV-Soc) cherry and their taxonomic
RL relationships within the species."; Arch. Virol. 148:2137-2153(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY184478; AAC62574.1; -; mRNA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002540; Pept_S30_Poty_P1.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001456; Peptidase_C6.
DR InterPro; IPR001592; Poty_coat.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00851; Peptidase_C6; 1.
DR Pfam; PF01577; Peptidase_S30; 1.
DR Pfam; PF00767; Poty_coat; 1.
DR Pfam; PF00680; RdRP_1; 1.
DR PRINTS; PR00966; NIAPOTYPTASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR Polyprotein.
KW SEQUENCE 3143 AA; 355410 MW; BEF5055767DBAE88 CRC64;

Query Match 57.3%; Score 47; DB 2; Length 3143;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PGKKVWLGSTSS 13
| | | | | : : : |
Db 311 PGKKFWVGFTNS 322

RESULT 21
Q858F7_9CAUD
ID Q858F7_9CAUD PRELIMINARY; PRT; 64 AA.
AC Q858F7;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE 18.
OS Enterobacteria phage epsilon15.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

OC P22-like viruses.
 OX NCBI_TaxID=215158;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Kropinski A.M., Billington S.J., Patrick A.N., Butts B.D.,
 RA Kovalyova I., McConnell M.R.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AY150271; AA006081.1; -; Genomic DNA.
 DR EMBL; AY150271; AA006081.1; -; Genomic DNA.
 SQ SEQUENCE 64 AA; 7010 MW; CE8DDBB6DEB5B88F CRC64;

 Query Match 56.1%; Score 46; DB 2; Length 64;
 Best Local Similarity 90.0%; Pred. No. 6.5;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 4 KKVWLGSTSS 13
 || |||||
 Db 34 KKGWLGSTSS 43

 RESULT 22
 Q6J9N4 ARATH
 ID Q6J9N4 ARATH PRELIMINARY; PRT; 157 AA.
 AC Q6J9N4;
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
 DT 05-JUL-2004, sequence version 1.
 DT 07-FEB-2006, entry version 10.
 DE Putative AP2/EREBP transcription factor.
 GN ORFNames=At2g40350;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Pan Y., Gong W., Liu D., Fu Q., Mei W.-Q., Song W.-Q., Ma L.-G.,
 RA Luo J.-C., Deng X.-W., Zhu Y.-X.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AY560891; AAT44958.1; -; mRNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2; 1.
 DR PRINTS; PD00367; ETHRSPLEPMNT.
 DR ProDom; PD001423; TFERF; 1.
 DR SMART; SM00380; AP2_1.
 DR PROSITE; PS1032; AP2_ERF; 1.
 SQ SEQUENCE 157 AA; 17758 MW; 9B5DF7BD39B42605 CRC64;

 Query Match 56.1%; Score 46; DB 2; Length 157;
 Best Local Similarity 61.5%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

 QY 3 GKKVWLGSTSSAY 15
 || |||||
 Db 88 GAKLWLGSTSSAY 100

 RESULT 23
 Q47TS8 THEFY
 ID Q47TS8 THEFY PRELIMINARY; PRT; 203 AA.
 AC Q47TS8;
 DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
 DT 13-SEP-2005, sequence version 1.

DT 07-FEB-2006, entry version 4.
 DE Hypothetical protein.
 GN OrderedLocustNames=Tfu 0098;
 OS Thermobifida fusca (strain YX).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
 OX NCBI_TaxID=269800;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RG US DOE Joint Genome Institute;
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hannon N., Istrati S., Pittluck S., Di Bartolo G., Chain P., Schmutz J.,
 RA Larimer F., Land M., Lykidis A., Richardson P.;
 RL "Complete sequence of Thermobifida fusca YX."
 RT Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; CP000088; AAZ54136.1; -; Genomic DNA.
 DR EMBL; CP000088; AAZ54136.1; -; Genomic DNA.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 203 AA; 21988 MW; E49349E55141ADD3 CRC64;

 Query Match 56.1%; Score 46; DB 2; Length 203;
 Best Local Similarity 72.7%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 3 GKKVWLGSTSS 13
 || |||||
 Db 163 GYKWLGDGTTSS 173

 RESULT 24
 Q7S5G5 NEUCR
 ID Q7S5G5 NEUCR PRELIMINARY; PRT; 827 AA.
 AC Q7S5G5;
 DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
 DT 15-DEC-2003, sequence version 1.
 DT 07-FEB-2006, entry version 12.
 DE Hypothetical protein.
 GN ORFNames=NCU06123.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
 RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
 RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
 RA Qui D., Ianakiev P., Bell-Pedersen D., Neilson M.A.,
 RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
 RA Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
 RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
 RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamayassellis M.,
 RA Maucelli E., Bielke C., Rudd S., Frishman D., Krystofova S.,
 RA Ramussen C., Metzberg R.L., Perkins D.D., Kroken S., Cogoni C.,
 RA Macino G., Catchside D.E.A., Li W., Pratt R.J., Osmani S.A.,
 RA DeSouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
 RA Varden O., Plamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
 RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
 RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.W.;
 RL "The genome sequence of the filamentous fungus Neurospora crassa."
 RL Nature 422:859-868(2003).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----


```

DR EMBL; AABX01000354; EAA30749.1; -; Genomic DNA.
DR GO; GO:0016832; F:aldehyde-lyase activity; IEA.
DR CO; CO:0016829; F:lyase activity; IEA.
DR GO; GO:0030976; F:thiamin pyrophosphate binding; IEA.
DR InterPro; IPR012109; Phosphoketolase.
DR Pfam; PF03894; XFP; 1.
DR PIRSF; PIRSF017245; Phosphoketolase; 1.
DR PROSITE; PS60002; PHOSPHOKETOLASE_1; UNKNOWN_1.
DR PROSITE; PS60003; PHOSPHOKETOLASE_2; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 827 AA; 61512 MW; A6D1CD869FB9B3E7 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 827;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GKKVWLGE 10
    |||||
DB 453 GKKVWLGE 460

RESULT 25
EX5B_ECOLI
ID EX5B_ECOLI STANDARD; PRT; 1180 AA.
AC P08394;
DT 01-AUG-1988, integrated into UniProtKB/Swiss-Prot.
DT 01-AUG-1988, sequence version 1.
DT 07-MAR-2006, entry version 54.
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5) (Exodeoxyribonuclease
DE V 135 kDa polypeptide).
GN Name=recB; Synonyms=roxA; OrderedLocusNames=b2820;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=V1000;
RX MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene.";
RL Nucleic Acids Res. 14:8573-8582(1986).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=V1000;
RX MEDLINE=20229837; PubMed=10766864; DOI=10.1074/jbc.275.16.12261;
RA Arnold D.A., Kowalczykowski S.C.;
RT "Facilitated loading of RecA protein is essential to recombination by
RT RecBCD enzyme.";
RL J. Biol. Chem. 275:12261-12265(2000).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=KL2 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1-11.
RX MEDLINE=87040734; PubMed=3534791;
RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli ptr gene
RT encoding protease III.";
RL Nucleic Acids Res. 14:7695-7703(1986).
RN [5]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1093-1180.
RX MEDLINE=87066730; PubMed=3537961;
RA Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of recD, the structural gene for the

```

```

RT alpha subunit of Exonuclease V of Escherichia coli.";
RL Nucleic Acids Res. 14:8583-8594(1986).
CC -!- FUNCTION: Required for efficient DNA repair; it catalyzes the
CC unwinding of double-stranded DNA and the cleavage of single-
CC stranded DNA and it stimulates local genetic recombination. All of
CC these activities require concomitant hydrolysis of ATP.
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -!- SUBUNIT: Consist of three subunits; recB, recC and recD.
CC -!- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL; X04581; CA28250.1; -; Genomic DNA.
EMBL; AF179304; AAD56369.1; -; Genomic DNA.
EMBL; U29581; AAB40467.1; -; Genomic DNA.
EMBL; U00096; AAC75859.1; -; Genomic DNA.
EMBL; X06227; CA229577.1; -; Genomic DNA.
EMBL; X04582; CA28252.1; -; Genomic DNA.
PIR; A25532; NCECK5.
PDB; 1W36; X-ray; B/E=1-1180.
GenomeReviews; U00096 GR; b2820.
EcoBASE; EB0817; -.
EcoGene; EGI0824; recB.
BioCyc; EcoCyc:EG10824-MONOMER; -.
GO; GO:0005515; F:protein binding; IPI.
InterPro; IPR004586; RecB.
InterPro; IPR000212; UvrD-helicase.
PANTHER; PTHR11070; UvrD-helicase; 2.
Pfam; PF00580; UvrD-helicase; 1.
TIGRFAMs; TIGR00609; recB; 1.
KW 3D-structure; ATP-binding; Complete proteome; DNA damage; DNA repair;
KW Endonuclease; Exonuclease; Helicase; Hydrolase; Nuclease;
KW Nucleotide-binding.
FT CHAIN 1 1180 Exodeoxyribonuclease V beta chain.
FT NP_BIND 23 30 /FTID=PRO_0000102046.
FT STRAND 3 4 ATP.
FT STRAND 6 7
FT HELIX 10 12
FT STRAND 16 17
FT STRAND 19 22
FT TURN 25 26
FT STRAND 27 27
FT HELIX 29 41
FT TURN 42 43
FT STRAND 45 49
FT STRAND 51 52
FT HELIX 56 58
FT STRAND 59 64
FT HELIX 66 88
FT STRAND 89 90
FT STRAND 94 94
FT HELIX 95 103
FT STRAND 105 105
FT HELIX 107 123
FT STRAND 125 128
FT HELIX 129 139
FT TURN 140 140
FT HELIX 141 144
FT TURN 145 145
FT STRAND 148 149
FT STRAND 151 152
FT HELIX 157 172
FT TURN 173 174
FT HELIX 177 186
FT STRAND 188 189
FT HELIX 190 197
FT TURN 198 201
FT STRAND 202 204
FT STRAND 207 210

```

FT TURN 214 215
FT TURN 218 234
FT TURN 235 236
FT STRAND 237 237
FT STRAND 240 241
FT STRAND 249 251
FT STRAND 253 253
FT TURN 254 254
FT HELIX 255 263
FT TURN 264 264
FT STRAND 267 268
FT STRAND 272 272
FT TURN 279 280
FT STRAND 281 282
FT HELIX 283 287
FT STRAND 309 309
FT TURN 311 311
FT HELIX 312 316
FT TURN 317 318
FT STRAND 319 319
FT STRAND 323 323
FT STRAND 324 346
FT TURN 347 347
FT STRAND 348 349
FT HELIX 351 363
FT STRAND 364 364
FT TURN 365 366
FT HELIX 367 377
FT STRAND 379 383
FT STRAND 385 385
FT HELIX 386 388
FT HELIX 391 401
FT TURN 402 403
FT TURN 405 406
FT STRAND 408 413
FT HELIX 415 417
FT TURN 421 424
FT STRAND 427 436
FT STRAND 440 442
FT STRAND 445 446
FT TURN 447 448
FT STRAND 449 449
FT HELIX 451 462
FT STRAND 463 466
FT TURN 467 468
FT STRAND 469 470
FT TURN 471 472
FT STRAND 473 473
FT STRAND 479 480
FT HELIX 482 484
FT TURN 485 486
FT STRAND 487 491
FT TURN 492 493
FT STRAND 494 496
FT STRAND 498 503
FT STRAND 506 507
FT TURN 511 512
FT HELIX 513 534
FT TURN 535 536
FT STRAND 538 542
FT TURN 543 544
FT STRAND 545 548
FT HELIX 551 553
FT STRAND 554 560
FT HELIX 561 572
FT TURN 573 575
FT STRAND 578 580
FT TURN 581 582
FT STRAND 586 586
FT HELIX 587 589
FT STRAND 590 591
FT HELIX 592 603
FT TURN 604 604

FT TURN 606 607
FT STRAND 608 608
FT HELIX 609 617
FT STRAND 618 618
FT HELIX 619 621

Query Match 56.1%; Score 46; DB 1; Length 1180;
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KKWLGESSAY 15
| | | | |
Db 1082 KSNWLGESSAY 1093

RESULT 26
Q31XG8 SHIDS
ID Q31XG8_SHIDS PRELIMINARY; PRT; 1180 AA.
AC Q31XG8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE ATP-dependent dsDNA/ssDNA exonuclease V subunit.
GN Name=recB; OrderedLocustNames=SBO_2710;
OS Shigella boydii serotype 4 (strain Sb227).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=300268;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16275786; DOI=10.1093/nar/gki954;
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,
Wang J., Xiong Z., Dong J., Xue Y., Xu X., Sun L., Chen S.,
Nie H., Peng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,
Qiang B., Hou Y., Yu J., Jin Q.;
RA "Genome dynamics and diversity of Shigella species, the etiologic
agents of bacillary dysentery.";
RL Nucleic Acids Res. 33:6445-6458(2005).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL; CP000036; ABB67240.1; -; Genomic_DNA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
KW Complete proteome; Exonuclease.
SQ SEQUENCE 1180 AA; 133973 MW; 184724E11BDDAAB CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KKWLGESSAY 15
| | | | |
Db 1082 KSNWLGESSAY 1093

RESULT 27
Q32CAL SHIDS
ID Q32CAL_SHIDS PRELIMINARY; PRT; 1180 AA.
AC Q32CAL;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE ATP-dependent dsDNA/ssDNA exonuclease V subunit.
GN Name=recB; OrderedLocustNames=SDY_3037;
OS Shigella dysenteriae serotype 1 (strain Sd197).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.

OX NCBI_TaxID=300267;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16275786; DOI=10.1093/nar/gki954;
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,
RA Wang J., Xiong Z., Dong J., Xue Y., Zhu Y., Xu X., Chen S.,
RA Nie H., Peng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,
RA Qiang B., Hou Y., Yu J., Jin Q.;
RT "Genome dynamics and diversity of Shigella species, the etiologic
RT agents of bacillary dysentery.";
RL Nucleic Acids Res. 33:6445-6458(2005).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL: CP000034; AB63054.1; -; Genomic DNA.
DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO: 0003677; F:DNA binding; IEA.
DR GO: 0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO: 0004327; F:exonuclease activity; IEA.
DR GO: 0006281; F:DNA repair; IEA.
KW Complete proteome; Exonuclease.
SQ SEQUENCE 1180 AA; 13394 MW; 119D5FF6F2D04EC2 CRC64;
Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 KKVWLGTSAY 15
Db 1082 KSNWLGDSAY 1093
RESULT 28
Q3YV40 SHISS
ID Q3YV40_SHISS PRELIMINARY; PRT; 1180 AA.
AC Q3YV40;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE DNA helicase ATP-dependent dcdna/ssDNA exonuclease V subunit.
GN Names=recB; OrderedLocNames=SSO_2977; ORFNames=SSO_2977;
OS Shigella sonnei (strain S8046).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=300269;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16275786; DOI=10.1093/nar/gki954;
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,
RA Wang J., Xiong Z., Dong J., Xue Y., Zhu Y., Xu X., Chen S.,
RA Nie H., Peng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,
RA Qiang B., Hou Y., Yu J., Jin Q.;
RT "Genome dynamics and diversity of Shigella species, the etiologic
RT agents of bacillary dysentery.";
RL Nucleic Acids Res. 33:6445-6458(2005).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL: CP000038; AA289572.1; -; Genomic DNA.
DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO: 0003677; F:DNA binding; IEA.
DR GO: 0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO: 0004527; F:exonuclease activity; IEA.
DR GO: 0004386; F:helicase activity; IEA.
DR GO: 0006281; F:DNA repair; IEA.
DR InterPro: IPR004586; RecB.
DR InterPro: IPR000212; UvrD-helicase.
DR Pfam: PF00580; UvrD-helicase; 1.
DR TIGRFAMs: TIGR00609; recB; 1.

KW Complete proteome; Exonuclease; Helicase.
SQ SEQUENCE 1180 AA; 133972 MW; 28B03E16E6341FEB CRC64;
Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 4 KKVWLGTSAY 15
Db 1082 KSNWLGDSAY 1093
RESULT 29
Q2MA17 ECOLI
ID Q2MA17_ECOLI PRELIMINARY; PRT; 1180 AA.
AC Q2MA17;
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Exonuclease V (RecBCD complex), beta subunit.
GN Name=recB;
OS Escherichia coli W3110.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=316407;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=81053692; PubMed=6159575;
RA Smith D.R., Calvo J.M.;
RT "Nucleotide sequence of the E coli gene coding for dihydrofolate
RT reductase.";
RL Nucleic Acids Res. 8:2255-2274 (1980).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RA Sekiguchi T., Ortega-Cesena J., Nosoh Y., Ohashi S., Tsuda K.,
RA Kanaya S.;
RT "DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of
RT Bacillus coagulans. Comparison with the enzymes of Saccharomyces
RT cerevisiae and Thermus thermophilus.";
RL Biochim. Biophys. Acta 867:36-44(1986).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RA Chen H., Sun Y., Stark T., Beattie W., Moses R.E.;
RT "Nucleotide sequence and deletion analysis of the polB gene of
RT Escherichia coli.";
RL DNA Cell Biol. 9:613-635(1990).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RA Smallshaw J.E., Kellin R.A.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase.";
RL Genetics (Life Sci. Adv.) 11:59-65(1992).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RA Hayashi K., Morooka N., Yamamoto Y., Fujita K., Isono K., Choi S.,
RA Ohtsubo E., Baba T., Wanner B.L., Mori H., Horiuchi T.;
RT "Highly accurate genome sequences of Escherichia coli K-12 strains
RT MG1655 and W3110.";
RL Mol. Syst. Biol. 0:0-0(2006).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX PubMed=16397293; DOI=10.1093/nar/gkj150;
RA Riley M., Abe T., Arnaud M.B., Berlyn M.K., Blattner F.R.,
RA Chaudhuri R.R., Glasner J.D., Horiuchi T., Keseler I.M., Kosuge T.,
RA Mori H., Perna N.T., Plunkett G. III, Rudd K.E., Serres M.H.,
RA Thomas G.H., Thomson N.R., Wishart D., Wanner B.L.;
RT "Escherichia coli K-12: a cooperatively developed annotation snapshot-

RT -2005.";
RL Nucleic Acids Res. 34:1-9(2006).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=97349980; PubMed=9205837; DOI=10.1093/dnares/4.2.91;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
Oshima T., Oyama S., Saico N., Sampei G., Satoh Y., Sivasundaram S.,
Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
Yamagata S., Horiuchi T.;
RC "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features";
RL DNA Res. 4:91-113(1997).
RN [8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=97251358; PubMed=9097040; DOI=10.1093/dnares/3.6.379;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
Yamamoto Y., Horiuchi T.;
RC "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map";
RL DNA Res. 3:379-392(1996).
RN [9]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=97251357; PubMed=9097039; DOI=10.1093/dnares/3.6.363;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RC "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377(1996).
RN [10]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=97094878; PubMed=8940112; DOI=10.1074/jbc.271.49.31145;
RA Arn E.A., Abelson J.N.;
RC "The 2'-5' RNA ligase of *Escherichia coli*. Purification, cloning, and
RT genomic disruption";
RL J. Biol. Chem. 271:31145-31153(1996).
RN [11]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=97061202; PubMed=8905232; DOI=10.1093/dnares/3.3.137;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RC "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map";
RL DNA Res. 3:137-155(1996).
RN [12]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RC "Systematic sequencing of the *Escherichia coli* genome: analysis of the
RT 2.4-4.1 min (110,917-193,643 bp) region";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [13]
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=K-12;
RX MEDLINE=94240115; PubMed=8183897;
RA Janosi L., Shimizu I., Kaj A.;
RC "Ribosome recycling factor (ribosome releasing factor) is essential
RT for bacterial growth";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).
RN [14]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=94124004; PubMed=7904973; DOI=10.1016/0378-1119(93)90470-N;
RA Allknetts R., Gerrard B.C., Court D., Dean M.C.;
RC "Cloning and organization of the abc and mdl genes of *Escherichia*
RT coli: relationship to eukaryotic multidrug resistance";
RL Gene 136:231-236(1993).
RN [15]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=94018640; PubMed=8412694;
RA van Heeswijk W.C., Rabenberg M., Westerhoff H.V., Kahn D.D.;
RC "The genes of the glutamine synthetase adenylylation cascade are not
RT regulated by nitrogen in *Escherichia coli*";
RL Mol. Microbiol. 9:443-458(1993).
RN [16]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93259920; PubMed=8387990;
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RC "Rhs elements of *Escherichia coli* K-12: complex composites of shared
RT and unique components that have different evolutionary histories";
RL J. Bacteriol. 175:2799-2808(1993).
RN [17]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93123180; PubMed=8419307;
RA Yamada M., Azaoka S., Saier M.H. Jr., Yamada Y.;
RC "Characterization of the gcd gene from *Escherichia coli* K-12 W3110 and
RT regulation of its expression";
RL J. Bacteriol. 175:568-571(1993).
RN [18]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93116053; PubMed=1474579;
RA Cormack R.S., Mackie G.A.;
RC "Structural requirements for the processing of *Escherichia coli* 5 S
RT ribosomal RNA by RNase E in vitro";
RL J. Mol. Biol. 228:1078-1090(1992).
RN [19]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93094132; PubMed=1459951;
RA Gervais F.G., Drapeau G.R.;
RC "Identification, cloning, and characterization of rcsF, a new
RT regulator gene for exopolysaccharide synthesis that suppresses the
RT division mutation *fts284* in *Escherichia coli* K-12";
RL J. Bacteriol. 174:8016-8022(1992).
RN [20]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RC "Identification and characterization of the *smbA* gene, a suppressor of
RT the *mukB* null mutant of *Escherichia coli*";
RL J. Bacteriol. 174:7517-7526(1992).
RN [21]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K-12;
RX MEDLINE=93011013; PubMed=1396599;
RA Condon C., Phillips J., Fu Z.Y., Squires C., Squires C.L.;
RC Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15
| | | | |
DB 1082 KSNWLGEDSSAY 1093

RESULT 30
Q8X6M9 ECO57 PRELIMINARY; PRT; 1180 AA.
AC Q8X6M9; 07AB60;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 24.
DE DNA helicase, ATP-dependent dsDNA/ssDNA exonuclease V subunit, ssDNA
DE endonuclease (DNA helicase RecB).
GN Name=recB; OrderedLocusNames=ECs3677, z4137;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Groetbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / Sakai / RIMD 050952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796; DOI=10.1093/dnares/8.1.11;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Ikida T., Takami H., Honda T., Sasaki G., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; BA000007; BAB37100.1; -; Genomic DNA.
DR EMBL; AE005174; AAG57931.1; -; Genomic DNA.
DR PIR; E91088; E91088.
DR PIR; G85933; G85933.
DR SMR; Q8X6M9; 1-1174.
DR BioCyc; ECOL83334-1-ECs3677-MONOMER; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0004003; F:ATP binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR00212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 2.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Complete proteome; Endonuclease; Exonuclease; Helicase.
SQ SEQUENCE 1180 AA; 134111 MW; D8CE1569E46F5F65 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15

DB 1082 KSNWLGEDSSAY 1093
| | | | |

RESULT 31
Q83JW0 SHIFL PRELIMINARY; PRT; 1180 AA.
AC Q83JW0_07UBS0;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-FEB-2005, sequence version 2.
DT 07-MAR-2006, entry version 19.
DE DNA helicase, ATP-dependent dsDNA/ssDNA exonuclease V subunit, ssDNA
DE endonuclease.
GN Name=recB; OrderedLocusNames=S3028, SF2831; ORFNames=S_3028;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
DOI=10.1128/JAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC
DR EMBL; AE005674; AAN44318.2; -; Genomic DNA.
DR EMBL; AE014073; AAP18143.1; -; Genomic DNA.
DR SMR; Q83JW0; 1-1174.
DR GenomeReviews; AE014073 GR; S3028.
DR GenomeReviews; AE005674 GR; SF2831.
DR BioCyc; SFLE198214:AA44318.1-MO-; -;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR00212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 2.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Complete proteome; Endonuclease; Exonuclease; Helicase.
SQ SEQUENCE 1180 AA; 134044 MW; A471BC694EC0D433 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15

```
Db      1082 KSNWLGEDSSAY 1093
|      ||||| |||||
RESULT 32
Q8FEB3 ECOL6 PRELIMINARY; PRT; 1183 AA.
AC Q8FEB3;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 21-FEB-2006, entry version 16.
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN Name=recB; ORFNames=c_3414;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=06:H1 / CF7073 / ATCC 700928 / UPEC;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; A2014075; AAN81859.1; -; Genomic_DNA.
DR SMR; Q8FEB3; 4-1177.
DR BioCyc; ECOL199310:C3414-MONOMER; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR PANTHER; PTHR11070; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRfams; TIGR00609; recB; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 1183 AA; 134424 MW; F11CB6919C16B8CA CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1183;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      4 KKWLGETSSAY 15
|      ||||| |||||
Db      1085 KSNWLGEDSSAY 1096
|      ||||| |||||
RESULT 33
Q8YKV0 ANASP PRELIMINARY; PRT; 1596 AA.
AC Q8YKV0;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE All7191 protein.
GN OrderedLocNames=all7191;
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
```

```
RX MEDLINE=21595285; PubMed=11759840; DOI=10.1093/dnares/8.5.205;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA000020; BAB78275.1; -; Genomic_DNA.
DR PIR; AG2501; AG2501.
DR BioCyc; NSP103690:ALL7191-MONOMER; -.
KW Complete proteome; Plasmid.
SQ SEQUENCE 1596 AA; 180618 MW; F812AD045822D289 CRC64;

Query Match 55.5%; Score 45.5; DB 2; Length 1596;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy      1 RPQ---KKWLGETSSA 14
|      ||| :|:|||||
Db      240 RPGLYKQIKWLGERKSQ 256
|      ||| :|:|||||
RESULT 34
Q3PI29 PARDE PRELIMINARY; PRT; 73 AA.
AC Q3PI29;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=PdenDRAFT_4242;
OS Paracoccus denitrificans PD1222.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=318586;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PD1222;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Paracoccus
denitrificans PD1222.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PD1222;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Paracoccus denitrificans
PD1222.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAIT01000002; EAM67635.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 73 AA; 8145 MW; 8C3DF87B487EAFB3 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 73;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 RPKKVLGETSSA 14
Db 20 RDGKVRGKCA 33

RESULT 35
Q41177_SYPN2
ID Q41177_SYPN2 PRELIMINARY; PRT; 134 AA.
AC Q41177;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE ORF134.
OS Synechococcus PCC7002 PR-6.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=34076;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=PCC 7002;
RA Akiyama H., Kanai S., Hirano M., Sugimoto M., Kiyohara M.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC
DR D13971; BAA03077.1; -; Genomic_DNA.
DR InterPro; IPR003435; Chaperonin_RcbX.
DR Pfam; PF02341; RcbX; 1.
SQ SEQUENCE 134 AA; 15269 MW; A11F42E96F424988 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 134;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PGKVLGETSSAY 15
Db 35 PGQVLGETSSAY 48

RESULT 36
DRE2H_ARATH
ID DRE2H_ARATH STANDARD; PRT; 164 AA.
AC Q9SI20;
DT 07-JUN-2004, integrated into UniProtKB/Swiss-Prot.
DT 07-JUN-2004, sequence version 2.
DT 07-MAR-2006, entry version 31.
DE Putative dehydration-responsive element-binding protein 2H (DREB2H
protein).
GN Name=DREB2H; OrderedLocusNames=At2g40350; ORFNames=T3G21.12;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197; DOI=10.1038/45471;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.-J., Rensing C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M.D., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana."
RL Nature 402:761-768(1999).
RN [2]
RN GENE FAMILY, AND FUNCTION.
RP MEDLINE=21656975; PubMed=11798174; DOI=10.1006/bbrc.2001.6299;
RA Sakuma Y., Liu Q., Dubouzet J.G., Abe H., Shinozaki K.,
```

```
RA Yamaguchi-Shinozaki K.;
RT "DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs,
RT transcription factors involved in dehydration- and cold-inducible gene
RT expression.";
RL Biochem. Biophys. Res. Commun. 290:998-1009(2002).
CC -!- FUNCTION: Putative transcriptional activator that binds
CC specifically to the DNA sequence 5'-(AG)CGAC-3' (By similarity).
CC -!- SUBCELLULAR LOCATION: Nucleus (probable).
CC -!- SIMILARITY: Contains 1 AP2/ERF DNA-binding domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction. May be the product of a pseudogene.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC
DR EMBL; AC007020; AAD25670.1; ALT_SEQ; Genomic_DNA.
DR PIR; D84828; D84828.
DR HSP; O80337; 2GCC.
DR GeneFarm; 3683; 373.
DR TAIR; At2g40350; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2; 1.
DR PRINTS; PR00367; ETRSPPELMNT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR PROSITE; PS51032; AP2_ERF; 1.
KW Activator; DNA-binding; Hypothetical protein; Nuclear protein;
KW Transcription; Transcription regulation.
FT CHAIN 1..164
FT Binding protein 2H.
FT /FTID=PRO_0000112541.
FT DNA_BIND 73..130 AP2/ERF.
FT MOTIF 10..50 Nuclear localization signal (Potential).
FT SEQUENCE 164 AA; 18493 MW; D816D272C9DFB8 CRC64;

Query Match 54.9%; Score 45; DB 1; Length 164;
Best Local Similarity 61.5%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GKKVLGETSSAY 15
Db 95 GAKVLGETSSAY 107

RESULT 37
Q5BK39_RAT
ID Q5BK39_RAT PRELIMINARY; PRT; 225 AA.
AC Q5BK39;
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.
DT 12-APR-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein LOC308990.
GN Name=LOC308990;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Thymus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Kleusner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thymus;
RG NIH MGC Project;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC091216; AAH91216.1; -; mRNA.
DR EMBL; BC091205; AAH91205.1; -; mRNA.
KW Hypothetical protein.
SQ SEQUENCE 225 AA; 24504 MW; 8158AD445C032754 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 225;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPQKKVWLGTTSSA 14
||| :|: |||
Db 76 RPQGLWIEPTSSA 89

RESULT 38
Q6P3A6_MOUSE PRELIMINARY; PRT; 225 AA.
ID Q6P3A6_MOUSE
AC Q6P3A6;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Hypothetical protein A1467606.
GN Name=A1467606;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland; DOI=10.1073/pnas.242603899;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Murzyn D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;

RG NIH MGC Project;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC064101; AAH64101.1; -; mRNA.
DR EMBL; ENSMUSG00000045165; Mus musculus.
KW Hypothetical protein.
SQ SEQUENCE 225 AA; 24489 MW; 8EB838636B92FC38 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 225;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPQKKVWLGTTSSA 14
||| :|: |||
Db 76 RPQGLWIEPTSSA 89

RESULT 39
Q8C708_MOUSE PRELIMINARY; PRT; 225 AA.
ID Q8C708_MOUSE
AC Q8C708;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE 0 day neonate kidney cDNA, RIKEN full-length enriched library,
DE clone: D630041B17 product: hypothetical protein, full insert sequence
DE (NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length
DE enriched library, clone: F630208N22 product: hypothetical protein, full
DE insert sequence) (Expressed sequence A1467606) (CRL-1722 L5178Y-R
DE cDNA, RIKEN full-length enriched library, clone: I730073I12
DE product: hypothetical protein, full insert sequence) (2 days neonate
DE thymic thymic cells cDNA, RIKEN full-length enriched library,
DE clone: E430005013 product: hypothetical protein, full insert sequence)
DE (NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length
DE enriched library, clone: F630035E07 product: hypothetical protein, full
DE insert sequence).
GN Name=A1467606;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44 (1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama T., Takahara S., Chihara T.,
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
Crawe M.L., Dalla E., Dalrymple B.P., De Bono B., Della Gatta G.,
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
Georgii-Hemmer P., Gingeras T.R., Gojobori T., Green R.E.,
Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
Hill D., Hummnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,
Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,

RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Oshiguro V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Sample C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugliura K., Sultana R., Takenaka Y., Taki K.,
RA Tamajima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamamichi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kal C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney, and Thymus; STRAIN=C57BL/6J, NOD, and DBA/2;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sample C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [5]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX MEDLINE=22085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaado I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stabli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baren G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guscinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Maehama J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [6]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [7]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-Format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [8]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [9]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=NOD;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN [10]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Mammary Gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.O., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Query Match          54.9%; Score 45; DB 2; Length 225;
Best Local Similarity 57.1%; Pred. No. 37; Mismatches 0; Gaps 0;
Matches 8; Conservative 3; Indels 3;

Qy 1 RPKKKVWLGETSSA 14
    ||| :|: ||||
Db 76 RFGGELWIEPTSSA 89

RESULT 40
Q8A4U3 BACTN      PRELIMINARY; PRT; 226 AA.
AC Q8A4U3;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN OrderedLocusNames=BT2504; ORFNames=BT_2504;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF015928; AA077611.1; -: Genomic_DNA.
DR Biocyc; BTHE226186:BT2504-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 226 AA; 26012 MW; 98FF9C9827FE6A71 CRC64;

Query Match          54.9%; Score 45; DB 2; Length 226;
Best Local Similarity 70.8%; Pred. No. 37; Mismatches 0; Gaps 0;
Matches 7; Conservative 2; Indels 1;

Qy 1 RPKKKVWLGE 10
    ||| :|: ||
Db 71 RFGERVWKGE 80

RESULT 41
Q3ITU3 NATPD      PRELIMINARY; PRT; 331 AA.
AC Q3ITU3;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Hypothetical protein.
GN OrderedLocusNames=NP0698A;
OS Natronomonas pharaonis (Strain DSM 2160 / ATCC 35678).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Natronomonas.
OX NCBI_TaxID=348780;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16169924; DOI=10.1101/gr.3952905;
RA Falb M., Pfeiffer F., Palm P., Rodewald K., Hickmann V., Tittor J.,
RA Oesterheld D.;
RT "Living with two extremes: conclusions from the genome sequence of
RT Natronomonas pharaonis.";
```

```
RL Genome Res. 15:1336-1343(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR936257; CAI48440.1; -: Genomic_DNA.
DR GenomeReviews; CR936257.GR; NF0698A.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 331 AA; 37126 MW; EC6E91B8FF0462AE CRC64;

Query Match          54.9%; Score 45; DB 2; Length 331;
Best Local Similarity 53.8%; Pred. No. 55;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RPKKKVWLGETSS 13
    ||| :|: |||
Db 65 RPRAYWLGNTET 77

RESULT 42
Q6FA89 ACIAD      PRELIMINARY; PRT; 335 AA.
AC Q6FA89;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN OrderedLocusNames=ACIAD2229;
OS Acinetobacter sp. (strain ADP1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=62977;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15514110; DOI=10.1093/nar/gkh910;
RA Barbe V., Vallent D., Fonknechten N., Kreimeyer A., Oztas S.,
RA Labarre L., Cruveiller S., Robert C., Duprat S., Wincker P.,
RA Ornstom L.N., Weissbach J., Marliere P., Cohen G.N., Medigue C.;
RT "Unique features revealed by the genome sequence of Acinetobacter sp.
RT ADP1, a versatile and naturally transformation competent bacterium.";
RL Nucleic Acids Res. 32:5766-5779(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR543861; CAG69024.1; -: Genomic_DNA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000312; Glyco trans.3.
DR Pfam; PF02885; Glycos trans 3N; 1.
DR Pfam; PF00591; Glycos_transf.3; 1.
DR ProDom; PD001864; Glyco_trans_3; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 335 AA; 38202 MW; EF189231512EDBAB CRC64;

Query Match          54.9%; Score 45; DB 2; Length 335;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KKVWLGETSSAY 15
    ||| :|: |||
Db 283 KSVWLGESSHVEY 294

RESULT 43
Q5XK86 XENLA      PRELIMINARY; PRT; 473 AA.
AC Q5XK86;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).;
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8155;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: BC083026; AAH83026.1; -; mRNA.
KW Hypothetical protein.
SQ SEQUENCE 473 AA; 51936 MW; 17EE6661F76CE267 CRC64;
Query Match 54.9%; Score 45; DB 2; Length 473;
Best Local Similarity 57.1%; Pred. No. 81;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 RPGRKWLGETSSA 14
:|||||:
DB 431 QPGKSIWLRESLA 444
RESULT 44
Q57MD5 SALCH
ID Q57MD5 SALCH PRELIMINARY; PRT; 562 AA.
AC Q57MD5;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DE Sugar specific PTS system, fructose-specific transport protein.
DE Name=frua; OrderedLocusNames=SC2220; ORFNames=SCH_2220;
OS Salmonella choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 9150 / SARB42;
RX PubMed=15531882; DOI=10.1038/ngl470;
RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,
RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozerky P., McLellan M.,
RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
RA Kohlberg S., Strong C., Du F., Carter J., Krenizki C., Layman D.,
RA Leonard S., Sun H., Fulton L., Naeh W., Miner T., Minx P.,
RA Delehaanty K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
RA Spieth J., Wilson R.K.;
RT "Comparison of genome degradation in Paratyphi A and Typhi, human-
RT restricted serovars of Salmonella enterica that cause typhoid";
RL Nat. Genet. 36:1268-1274(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: CP000026; AAV76647.1; -; Genomic DNA.

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495; DOI=10.1093/nar/gki297;
RA Chiu C.-H., Tang F., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
RA Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
RT highly invasive and resistant zoonotic pathogen";
RL Nucleic Acids Res. 33:1690-1698(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AE017220; AAX66126.1; -; Genomic DNA.
GO GO:0016021; C:integral to membrane; IEA.
GO GO:0016020; C:membrane; IEA.
GO GO:0008982; F:protein-N(P)-phosphohistidine-sugar phospho. .; IEA.
GO GO:0005351; F:sugar porter activity; IEA.
GO GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.
DR InterPro: IPR013011; PTS_EIIB_2.
DR InterPro: IPR003352; PTS_EIIC.
DR InterPro: IPR013014; PTS_EIIC_2.
DR InterPro: IPR003353; PTS_IIB fruc.
DR Pfam: PF02378; PTS_EIIC_1.
DR Pfam: PF02379; PTS_IIB fruc; 1.
DR TIGRFAMs: TIGR00829; FRU; 1.
DR TIGRFAMs: TIGR01427; PTS_IIC fructo; 1.
DR PROSITE: PS51099; PTS_EIIB TYPE 2; 1.
DR PROSITE: PS51104; PTS_EIIC TYPE 2; 1.
KW Complete proteome.
SQ SEQUENCE 562 AA; 57370 MW; 15BA06985E935A74 CRC64;
Query Match 54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred. No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 3 GKXVWLGITSSA 14
|||||:
DB 59 GKXVWLGITGRA 70
RESULT 45
Q5PE41 SALPA
ID Q5PE41 SALPA PRELIMINARY; PRT; 562 AA.
AC Q5PE41;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE PTS system, fructose-specific IIBC component.
GN Name=frua; OrderedLocusNames=SPA0647;
OS Salmonella paratyphi-a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=54388;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 9150 / SARB42;
RX PubMed=15531882; DOI=10.1038/ngl470;
RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,
RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozerky P., McLellan M.,
RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
RA Kohlberg S., Strong C., Du F., Carter J., Krenizki C., Layman D.,
RA Leonard S., Sun H., Fulton L., Naeh W., Miner T., Minx P.,
RA Delehaanty K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
RA Spieth J., Wilson R.K.;
RT "Comparison of genome degradation in Paratyphi A and Typhi, human-
RT restricted serovars of Salmonella enterica that cause typhoid";
RL Nat. Genet. 36:1268-1274(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: CP000026; AAV76647.1; -; Genomic DNA.

DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . . ; IEA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . . ; IEA.
DR InterPro; IPR013011; PTS_EIIB_2.
DR InterPro; IPR003352; PTS_EIIC.
DR InterPro; IPR013014; PTS_EIIC_2.
DR InterPro; IPR003353; PTS_IIB fruc.
DR InterPro; IPR006327; PTS_IIC fruc.
DR Pfam; PF02378; PTS_EIIC; 1.
DR Pfam; PF02379; PTS_IIB fruc; 1.
DR TIGRFAMs; TIGR00829; FRU; 1.
DR TIGRFAMs; TIGR01427; PTS_IIC fructo; 1.
DR PROSITE; PS1099; PTS_EIIB TYPE 2; 1.
DR PROSITE; PS1104; PTS_EIIC TYPE 2; 1.
KW Complete proteome.
SQ SEQUENCE 562 AA; 57289 MW; B0302A85DDDBE865 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred.No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14
|||||: |
Db 59 GKKVWLGDIGRA 70

RESULT 46
Q8ZNK5 SALTY PRELIMINARY; PRT; 562 AA.
AC Q8ZNK5;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-FEB-2002, sequence version 1.
DT 07-FEB-2006, entry version 20.
GN Fructose-specific transport protein (EC 2.7.1.69).
GN Name=fruA; OrderedLocusNames=STM2204;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RT Nature 413:852-856 (2001).
RL Nature 413:852-856 (2001).
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AE007998; AL211108.1; -; Genomic_DNA.
DR BioCyc; STVP9287:STM2204-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . . ; IEA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . . ; IEA.
DR InterPro; IPR013011; PTS_EIIB_2.
DR InterPro; IPR003352; PTS_EIIC.
DR InterPro; IPR013014; PTS_EIIC_2.
DR InterPro; IPR003353; PTS_IIB fruc.
DR InterPro; IPR006327; PTS_IIC fruc.
DR Pfam; PF02378; PTS_EIIC; 1.
DR Pfam; PF02379; PTS_IIB fruc; 1.
DR TIGRFAMs; TIGR00829; FRU; 1.

DR TIGRFAMs; TIGR01427; PTS_IIC fructo; 1.
DR PROSITE; PS1099; PTS_EIIB TYPE 2; 1.
DR PROSITE; PS1104; PTS_EIIC TYPE 2; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 562 AA; 57375 MW; 3D6EE64663F8AA2C CRC64;

Query Match 54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred.No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14
|||||: |
Db 59 GKKVWLGDIGRA 70

RESULT 47
Q8Z592 SALTI PRELIMINARY; PRT; 562 AA.
AC Q8Z592; Q7CB39;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-MAR-2006, entry version 24.
DE PTS system, fructose-specific IIBC component (Fructose-specific IIBC component of PTS system).
GN Name=fruA; OrderedLocusNames=STY2439, t0651;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Conerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
RT Nature 413:848-852 (2001).
RL Nature 413:848-852 (2001).
[2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RX DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";
RT J. Bacteriol. 185:2330-2337 (2003).
RL J. Bacteriol. 185:2330-2337 (2003).
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AL627273; CAD02586.1; -; Genomic_DNA.
DR EMBL; AE014613; AAO68352.1; -; Genomic_DNA.
DR GenomeReviews; AL513382 GR; STY2439.
DR GenomeReviews; AE014613 GR; t0651.
DR BioCyc; SNT029461:T0651-MONOMER; -.
DR BioCyc; SNT90370:STY2439-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . . ; IEA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . . ; IEA.
DR InterPro; IPR013011; PTS_EIIB_2.
DR InterPro; IPR003352; PTS_EIIC.
DR InterPro; IPR013014; PTS_EIIC_2.
DR InterPro; IPR003353; PTS_IIB fruc.

```
DR InterPro; IPR006327; PTS_IIC_fruc.
DR Pfam; PF02378; PTS_IIC_1.
DR Pfam; PF02379; PTS_IIB_fruc; 1.
DR TIGRFAMs; TIGR00829; FRU; 1.
DR TIGRFAMs; TIGR01427; PTS_IIC_fructo; 1.
DR PROSITE; PS1099; PTS_IIB_TYPE_2; 1.
DR PROSITE; PS1104; PTS_IIC_TYPE_2; 1.
KW Complete proteome.
SQ SEQUENCE 562 AA; 57289 MW; B0302A85DDDB865 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred. No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGTSAA 14
DB 59 GKKVWLDIGRA 70
|||||:|

RESULT 48
Q5SE54_DICDI PRELIMINARY; PRT; 962 AA.
AC Q5SE54;
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
GN ORFNames=DD80190225;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RX PubMed=15875012; DOI=10.1038/nature03481;
RA Eichinger L., Pachbat J.A., Gloeckner G., Rajandream M.A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankal A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E.J., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.M.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D.M., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Buchrieser C., Wardrop A., Felder M., Thangavelu M., Saito T.,
RA Knights A., Loulseghe H., Mungall K.L., Oliver K., Price C.,
RA Quail M.A., Urushihara H., Hernandez J., Rabinowitsch E., Steffen D.,
RA Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M.N., Spiegler S.,
RA Tivey A., Sugano S., White B., Walker D., Woodward J.R., Winckler T.,
RA Tanaka Y., Shaulsky G., Schleicher M., Weinstein G.M., Rosenthal A.,
RA Cox E.C., Chisholm R.L., Gibbs R.A., Loomis W.F., Platzer M.,
RA Kay R.C., Williams J.G., Dear P.H., Nogel A.A., Barrell B.G.,
RA Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 435:43-57(2005).
CC CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAFI01000011; EAL72044.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD0000018; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00682; WD_REPEATS_2; 4.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.

SQ SEQUENCE 962 AA; 105955 MW; 1545D71B703CID96 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 962;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGTSAY 15
DB 711 GRAVWLGQTGNIF 723
|||||:|

RESULT 49
Q3HF42_TRIER PRELIMINARY; PRT; 1869 AA.
AC Q3HF42;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Alpha-2-macroglobulin, N-terminal:2-oxo acid dehydrogenase, lipoyl-
DE binding site.
GN ORFNames=teryDRAFT 2818;
OS Trichodesmium erythraeum IMS101.
OC Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX NCBI_TaxID=203124;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IMS101;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Trichodesmium
RT erythraeum IMS101.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IMS101;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Trichodesmium erythraeum
RT IMS101.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABK04000004; EAO28072.1; -; Genomic_DNA.
DR GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR InterPro; IPR002890; A2M_N.
DR InterPro; IPR011625; A2M_N_2.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01835; A2M_N; 1.
DR Pfam; PF07703; A2M_N_2; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SEQUENCE 1869 AA; 209056 MW; 9733B25719B73599 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 1869;
Best Local Similarity 60.0%; Pred. No. 3.5e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 1 RPKKKVWLGTSAY 15
DB 598 QPGEKVVWL--TCAAY 610
|||||:|
```

```
RESULT 50
Q36TJ3_MARHY
ID Q36TJ3_MARHY PRELIMINARY; PRT; 197 AA.
AC Q36TJ3;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Phosphonate metabolism.
GN ORFNames=MaquDRAFT_2498;
OS Marinobacter aquaeolei VT8.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Marinobacter.
OX NCBI_TaxID=351348;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Marinobacter aquaeolei
RT VT8.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Marinobacter aquaeolei VT8.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
DR EMBL; AALG01000003; EAP00158.1; -; Genomic DNA.
DR GO; GO:0015716; P:phosphonate transport; IEA.
SQ SEQUENCE 197 AA; 21341 MW; 8BA55E15F3B7C259 CRC64;

Query Match 53.7%; Score 44; DB 2; Length 197;
Best Local Similarity 53.8%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSA 14
DB 42 PGESVWLADTDGA 54

Search completed: June 5, 2006, 12:51:58
Job time : 175.507 secs
```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 104.384 Seconds

(without alignments)
65.702 Million cell updates/sec

Title: US-10-645-659A-8

Perfect score: 82

Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

9: Geneseqp2005s:*

10: Geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	8	ADR88214 Human hep
2	82	100.0	15	8	ADT78181 Functiona
3	82	100.0	15	9	ADU70829 Human hep
4	82	100.0	15	9	AEA42430 Human hep
5	82	100.0	386	8	ADR88207 Human mat
6	82	100.0	386	8	ADT78174 45kDa sub
7	82	100.0	386	9	ADY27057 Heparanas
8	82	100.0	386	9	ADZ18995 Human hep
9	82	100.0	386	9	AEA42423 Human mat
10	82	100.0	460	9	ADY27061 Heparanas
11	82	100.0	486	9	AE887589 Human hep
12	82	100.0	492	9	ADZ18996 Hep106 co
13	82	100.0	493	9	AE887562 Human hep
14	82	100.0	495	9	ADZ18999 Hep109 co
15	82	100.0	497	9	AE887587 Human hep
16	82	100.0	501	9	ADZ19000 HepG3 co
17	82	100.0	507	9	ADY219005 HepG56 co
18	82	100.0	508	9	ADY27058 Human ina
19	82	100.0	526	9	ADZ19006 HepHyalur
20	82	100.0	527	5	ABW07815 Chicken s
21	82	100.0	527	7	ABW02018 Chimeric
22	82	100.0	527	8	ADO63826 Chimeric
23	82	100.0	527	9	ADZ19004 HepG54 co

97	68	82.9	15	9	ADU70898	Human hep	170	41	50.0	286	3	AAG48720	Aag48720 Arabidops
98	67	81.7	15	9	ADU71196	Human hep	171	41	50.0	286	3	AAG11224	Aag11224 Arabidops
99	64	78.0	15	9	ADU70961	Human hep	172	41	50.0	291	3	AAG48719	Aag48719 Arabidops
100	62	75.6	15	9	ADU71197	Human hep	173	41	50.0	291	3	AAG11223	Aag11223 Arabidops
101	58	70.7	15	9	ADU71198	Human hep	174	41	50.0	298	3	AAG11222	Aag11222 Arabidops
102	58	70.7	528	5	AAE18327	Human hep	175	41	50.0	298	3	AAG48718	Aag48718 Arabidops
103	58	70.7	538	4	AAE18327	Human hep	176	41	50.0	303	4	ABB63233	Abb63233 Drosophil
104	58	70.7	582	4	AAE18326	Human hep	177	41	50.0	431	4	AAB20603	Aau20603 Human sec
105	58	70.7	592	4	AAE18326	Human hep	178	41	50.0	515	8	ADX92843	Adx92843 Plant ful
106	58	70.7	592	4	AAE18326	Human hep	179	41	50.0	538	6	ABM68256	Abm68256 Photorhab
107	58	70.7	592	4	AAE18326	Human hep	180	41	50.0	648	7	ADD45479	Add45479 Rat Prote
108	58	70.7	592	4	AAE18326	Human hep	181	41	50.0	648	7	ADDS7446	Ades7446 Rat Prote
109	54	65.9	9	9	ADU70758	Human hep	182	41	50.0	648	7	ABE18192	Aeb18192 Norway ra
110	54	65.9	9	9	ADU70693	Human hep	183	41	50.0	1146	4	AAE00961	Aae00961 Chicken l
111	54	65.9	15	9	ADU70921	Human hep	184	41	50.0	1194	6	ABU18653	Abu18653 Protein e
112	52	63.4	15	9	ADU71044	Human hep	185	40	48.8	45	9	AEC20423	Aec20423 FIV trime
113	52	63.4	651	9	ABE18194	Dog beta-	186	40	48.8	145	6	ABO44374	AbO44374 Corn ster
114	50	61.0	651	9	ABE18193	Cat beta-	187	40	48.8	237	9	ADY53112	Ady53112 Acropora
115	49	59.8	9	9	ADU70795	Human hep	188	40	48.8	237	9	ADY53113	Ady53113 Acropora
116	49	59.8	9	9	ADU70543	Human hep	189	40	48.8	340	4	ABE71784	Abb71784 Drosophil
117	48	58.5	9	9	ADU70759	Human hep	190	40	48.8	433	3	ABM34745	Abm34745 Arabidops
118	48	58.5	9	9	ADU70387	Human hep	191	40	48.8	439	6	ABM67775	Abm67775 Photorhab
119	48	58.5	9	9	ADU70577	Human hep	192	40	48.8	498	9	ADX39546	Adx39546 HIV Gag p
120	48	58.5	15	9	ADU71195	Human hep	193	40	48.8	505	2	AAK48206	Aar48206 A.thalian
121	46	56.1	9	9	ADU70600	Human hep	194	40	48.8	505	3	ABM34744	Aar34744 Arabidops
122	46	56.1	17	9	ADZ18993	Synthetic	195	40	48.8	505	9	AEC59834	Aec59834 Cinnamate
123	45	54.9	463	8	ADZ25621	Aspergill	196	40	48.8	505	10	AEF28197	Aef28197 Lead Cere
124	45	54.9	488	4	ABM31469	Amino aci	197	40	48.8	533	3	ABM34743	Aag34743 Arabidops
125	45	54.9	488	4	ABM31472	Amino aci	198	40	48.8	549	8	ABM84770	Abm84770 Human dia
126	45	54.9	488	4	ABM31470	Amino aci	199	40	48.8	572	8	ABM84769	Abm84769 Human dia
127	45	54.9	648	4	AAE02444	Murine be	200	40	48.8	572	8	ABM84768	Abm84768 Human dia
128	45	54.9	648	7	ADF28924	Murine be	201	40	48.8	589	8	ABM84772	Abm84772 Human dia
129	45	54.9	648	9	ABE18191	Mouse bet	202	40	48.8	600	7	ADG76363	Adg76363 Human Inc
130	44	53.7	90	5	ABP26310	Streptoco	203	40	48.8	600	8	ABM84771	Abm84771 Human dia
131	44	53.7	364	4	AAE13810	Peptide #	204	40	48.8	613	2	AAW93823	Aaw93823 E. coli G
132	44	53.7	364	4	ABM32748	Peptide #	205	40	48.8	613	2	AAW93828	Aaw93828 Human GUS
133	44	53.7	364	4	ABM32748	Peptide #	206	40	48.8	613	3	ABM28407	Aeb28407 Escherich
134	44	53.7	364	4	ABM26210	Peptide #	207	40	48.8	628	9	ABE18189	Aeb18189 Nematode
135	44	53.7	364	4	ABM27585	Human pep	208	40	48.8	633	4	ABM62276	Abm62276 Mutant he
136	44	53.7	364	4	ABM18230	Protein #	209	40	48.8	633	4	ABM62271	Abm62271 Heavy cha
137	44	53.7	364	4	AAE65943	Human bon	210	40	48.8	640	2	AY27210	Aay27210 Amino aci
138	44	53.7	364	4	AAE53560	Human bra	211	40	48.8	648	9	ABE18195	Aeb18195 African g
139	44	53.7	364	4	ABG47603	Human liv	212	40	48.8	651	4	AAE02443	Aae02443 Human bet
140	44	53.7	364	5	ABG35580	Human pep	213	40	48.8	651	7	ADD45481	Add45481 Human pro
141	44	53.7	377	4	ABU52692	Human dif	214	40	48.8	651	7	ADDE57448	Ades7448 Human pro
142	44	53.7	383	4	ABG04768	Novel hum	215	40	48.8	651	8	ADP12392	Adp12392 Protein e
143	44	53.7	438	4	ABU52691	Different	216	40	48.8	651	8	ADQ98908	Adq98908 Antagonis
144	44	53.7	438	8	ADO44021	Amino aci	217	40	48.8	651	9	ADX06032	Adx06032 Cyclin-de
145	44	53.7	438	9	ABE11877	Testis sp	218	40	48.8	651	9	ABE18196	Aeb18196 Human bet
146	44	53.7	463	3	ABM41709	Human ORF	219	40	48.8	651	9	AED74652	Aed74652 Human pla
147	42	51.2	67	7	ADC95420	E. faeciu	220	40	48.8	716	8	ADS44698	Ads44698 Bacterial
148	42	51.2	121	5	ABP02395	Human ORF	221	40	48.8	716	10	AEF94558	Aef94558 Bacillus
149	42	51.2	207	4	ABG03263	Novel hum	222	40	48.8	722	6	ABE33322	Aeb33322 L. mexica
150	42	51.2	250	4	ABG79652	Corynebac	223	40	48.8	722	8	ADP47503	Adp47503 Human bet
151	42	51.2	250	4	AAE04864	Corynebac	224	40	48.8	722	8	ADJ58609	Adj58609 Fusion pr
152	42	51.2	250	4	AAU71880	C. glutam	225	40	48.8	722	10	AEF05094	Aef05094 Beta-gluc
153	42	51.2	250	4	AAU71880	C. glutami	226	40	48.8	799	5	ABP73572	Abp73572 Candida a
154	42	51.2	250	5	ABM18986	C. glutam	227	40	48.8	909	2	AAE50092	Aar50092 Humanised
155	42	51.2	250	5	ABM18986	C. glutam	228	40	48.8	7067	6	AAE35265	Aae35265 Human P45
156	42	51.2	250	6	ABP98290	Amino aci	229	39.5	48.2	594	8	ADQ36921	Adq36921 Cell prol
157	42	51.2	656	9	ABJ26836	Thymidyla	230	39.5	48.2	594	4	AAU08489	Aau08489 Human VMG
158	42	51.2	656	9	ABJ26837	Thymidyla	231	39.5	48.2	676	5	AAU76161	Aau76161 Bacillus
159	42	51.2	250	6	ABR55231	Amino aci	232	39.5	48.2	676	10	AEF68907	Aef68907 B. stearo
160	42	51.2	250	6	ABR55231	Amino aci	233	39	47.6	52	4	AAE82471	Aam82471 Human imm
161	42	51.2	250	6	ABR55231	Amino aci	234	39	47.6	94	5	ABG93137	Abg93137 S. cerevi
162	42	51.2	656	9	AED25678	Sunflower	235	39	47.6	137	8	ADS16684	Ads16684 Bartonell
163	42	51.2	1420	5	ABP73769	Candida a	236	39	47.6	258	7	ADH87709	Adh87709 Enterococ
164	41	50.0	9	9	ADU70491	Human hep	237	39	47.6	280	7	ADJ80150	Adj80150 Novel hum
165	41	50.0	15	9	ADU71109	Human hep	238	39	47.6	280	9	AEC20252	Aec20252 Human nuc
166	41	50.0	15	9	ADU71042	Human hep	239	39	47.6	354	4	ABE60901	Aeb60901 Drosophil
167	41	50.0	121	9	ABE39609	L. pneumo	240	39	47.6	357	2	AAE77394	Aar77394 Fragment
168	41	50.0	125	9	ABE36186	L. pneumo	241	39	47.6	362	8	ADS23337	Ads23337 Bacterial
169	41	50.0	202	7	ABO84289	Pseudomon	242	39	47.6	373	7	ADJ38293	Adj38293 A. cellul

243	39	47.6	375	6	ABP73019	Abp73019 Amino aci	316	38	46.3	410	3	AAG47939	Aag47939 Arabidops
244	39	47.6	389	5	AAM49427	Aam49427 Penicilli	317	38	46.3	431	5	ABB90809	Abb90809 Herbicida
245	39	47.6	403	7	ADC32901	Adc32901 Human nov	318	38	46.3	432	3	AAG47938	Aag47938 Arabidops
246	39	47.6	435	2	AAV19792	Aav19792 B. burgdo	319	38	46.3	451	3	AAG20843	Aag20843 Arabidops
247	39	47.6	454	2	AAV19791	Aav19791 B. burgdo	320	38	46.3	455	3	AAG20842	Aag20842 Arabidops
248	39	47.6	473	8	ADS44165	Ads44165 Bacterial	321	38	46.3	488	4	AAB31471	Aab31471 Amino aci
249	39	47.6	478	3	AAG29660	Aag29660 Arabidops	322	38	46.3	634	6	ABU18213	Abu18213 Protein e
250	39	47.6	481	3	AAG29659	Aag29659 Arabidops	323	38	46.3	726	6	ABR53854	Abri53854 Protein e
251	39	47.6	491	6	ABU20159	Abu20159 Protein e	324	38	46.3	726	6	ADK64928	Adk64928 Disease t
252	39	47.6	620	8	ADS25748	Ads25748 Bacterial	325	38	46.3	726	6	ABR53854	Abri53854 Protein e
253	39	47.6	623	8	ADS25902	Ads25902 Bacterial	326	38	46.3	2307	4	ABB54333	Abb54333 Drosophil
254	39	47.6	647	8	ABU40705	Abu40705 Protein e	327	38	46.3	3290	6	ADA34199	Ada34199 Acinetoba
255	39	47.6	649	8	ADS22340	Ads22340 Bacterial	327	37.5	45.7	328	6	ABU20472	Abu20472 Protein e
256	39	47.6	663	7	AAW49872	Aaw49872 Thermotog	328	37	45.1	38	4	ABG00367	Abg00367 Human nov
257	39	47.6	663	7	ADC26952	Adc26952 Thermotog	329	37	45.1	55	4	AAU20674	Aau20674 Human nov
258	39	47.6	663	7	ADF04165	Adf04165 Bacterial	330	37	45.1	78	2	AAV35853	Aav35853 Chlamydia
259	39	47.6	663	7	ADS93948	Ads93948 T. mariti	331	37	45.1	96	5	ABP42516	Abp42516 Human ova
260	39	47.6	666	2	AAW34992	Aaw34992 Thermotog	332	37	45.1	99	4	AAM84777	Aam84777 Human imm
261	39	47.6	680	2	AAW34564	Aaw34564 Thermotog	333	37	45.1	103	8	ADL05874	Adl05874 M. catarr
262	39	47.6	680	2	AAW35005	Aaw35005 Thermotog	334	37	45.1	112	4	ABG16641	Abg16641 Novel hum
263	39	47.6	680	2	AAW49868	Aaw49868 Thermotog	335	37	45.1	156	6	ABU29335	Abu29335 Protein e
264	39	47.6	680	7	ADC26916	Adc26916 Thermotog	336	37	45.1	157	4	ABE64397	Abi64397 Amino aci
265	39	47.6	680	7	ADB93812	Adb93812 T. mariti	337	37	45.1	158	7	ADH88103	Adh88103 Enterococ
266	39	47.6	699	3	AAW34445	Aaw34445 Candida a	338	37	45.1	174	8	ADP84553	Adp84553 Human bre
267	39	47.6	699	5	ABP73798	Abp73798 Candida a	339	37	45.1	192	4	ABB65518	Abb65518 Drosophil
268	39	47.6	762	6	ABP73022	Abp73022 Amino aci	340	37	45.1	224	2	AAV37471	Aav37471 Protein i
269	39	47.6	762	7	ADJ38291	Adj38291 A. celluli	341	37	45.1	225	3	AAG13478	Aag13478 Arabidops
270	39	47.6	941	2	AAW07478	Aaw07478 Cellulase	342	37	45.1	247	3	AAG13477	Aag13477 Arabidops
271	39	47.6	941	2	AAW77395	Aaw77395 Full leng	343	37	45.1	253	3	AAG43712	Aag43712 Arabidops
272	39	47.6	978	4	ABG13924	Abg13924 Novel hum	344	37	45.1	257	3	AAG43711	Aag43711 Arabidops
273	38	46.3	33	6	ABP79666	Abp79666 N. gonorr	345	37	45.1	279	3	AAG43710	Aag43710 Arabidops
274	38	46.3	65	5	ADH32561	Adh32561 Yeast smo	346	37	45.1	285	6	ADA35177	Ada35177 Acinetoba
275	38	46.3	97	4	ABG10472	Abg10472 Novel hum	347	37	45.1	316	9	ABE40822	Abi40822 L. pneumo
276	38	46.3	119	4	ADH80700	Adh80700 Human pol	348	37	45.1	319	9	ABE37510	Abi37510 L. pneumo
277	38	46.3	120	4	AAU14382	Aau14382 Human nov	349	37	45.1	327	9	ABE48805	Abi48805 Streptomy
278	38	46.3	120	10	ABE24056	Abe24056 Novel hum	350	37	45.1	342	6	ABU35760	Abu35760 Protein e
279	38	46.3	128	8	ADT59176	Adt59176 Plant pol	351	37	45.1	374	9	ABM96740	Abm96740 M. xanthu
280	38	46.3	137	5	ABW47884	Abw47884 p15 Ink4b	352	37	45.1	377	6	ABP76792	Abp76792 N. gonorr
281	38	46.3	138	2	AAW85115	Aaw85115 Cell-cycl	353	37	45.1	377	6	ABP81027	Abp81027 N. gonorr
282	38	46.3	138	3	AAW88361	Aaw88361 Human cel	354	37	45.1	395	8	ADS29563	Ads29563 Bacterial
283	38	46.3	138	5	AAU74701	Aau74701 Human cel	355	37	45.1	408	7	ADG66862	Adg66862 Klebsiell
284	38	46.3	160	8	ADV89860	Adv89860 Streptoco	356	37	45.1	415	4	AAE00417	Aae00417 Lycopersi
285	38	46.3	160	8	ADV82606	Adv82606 Streptoco	357	37	45.1	420	7	ADG62943	Adg62943 Rat Prote
286	38	46.3	160	8	ADV83250	Adv83250 Streptoco	358	37	45.1	420	7	ADG62949	Adg62949 Rat Prote
287	38	46.3	160	8	ADV81113	Adv81113 Streptoco	359	37	45.1	420	7	ADG62946	Adg62946 Rat Prote
288	38	46.3	160	8	ADV82665	Adv82665 Streptoco	360	37	45.1	420	7	ADG62949	Adg62949 Rat Prote
289	38	46.3	196	8	ADX89925	Adx89925 Plant ful	361	37	45.1	420	7	ADG62949	Adg62949 Rat Prote
290	38	46.3	207	8	ADX94844	Adx94844 Plant ful	362	37	45.1	430	7	ADG63970	Adg63970 Disease t
291	38	46.3	213	8	ADS26087	Ads26087 Bacterial	363	37	45.1	430	8	ADN19013	Adn19013 Bacterial
292	38	46.3	215	8	ADS25518	Ads25518 Bacterial	364	37	45.1	437	7	ADB80185	Adb80185 Mycobacte
293	38	46.3	215	8	ADS22628	Ads22628 Bacterial	365	37	45.1	442	6	ABU38438	Abu38438 Protein e
294	38	46.3	215	8	ADS25199	Ads25199 Bacterial	366	37	45.1	458	8	ADK48652	Adk48652 Streptoco
295	38	46.3	241	8	ADX93484	Adx93484 Plant ful	367	37	45.1	463	8	ADK48652	Adk48652 Streptoco
296	38	46.3	247	2	AAW66291	Aaw66291 Mycobacte	368	37	45.1	463	9	AEA58741	Aea58741 Novel S.
297	38	46.3	247	2	AAW66291	Aaw66291 Mycobacte	369	37	45.1	463	9	AEA58741	Aea58741 Streptoco
298	38	46.3	247	2	AAW63899	Aaw63899 M. bovis	370	37	45.1	479	8	ADT56538	Adt56538 Plant pol
299	38	46.3	247	2	AAW40809	Aaw40809 M. bovis	371	37	45.1	485	7	ABO77109	Abi77109 Pseudomon
300	38	46.3	247	6	ABU36643	Abu36643 Protein e	372	37	45.1	518	6	ABU34324	Abu34324 Protein e
301	38	46.3	247	6	ABU34433	Abu34433 Protein e	373	37	45.1	521	8	ADK92593	Adk92593 Plant ful
302	38	46.3	250	8	ADX76897	Adx76897 Plant ful	374	37	45.1	544	2	AAW82213	Aaw82213 Talaromyc
303	38	46.3	260	8	ADX88830	Adx88830 Plant ful	375	37	45.1	552	8	ADT57791	Adt57791 Plant pol
304	38	46.3	260	8	ADX92376	Adx92376 Plant ful	376	37	45.1	554	8	ABM84754	Abm84754 Human dia
305	38	46.3	324	3	AAG47940	Aag47940 Arabidops	377	37	45.1	557	8	ADK75975	Adk75975 Plant ful
306	38	46.3	324	3	ABW61517	Abw61517 Drosophil	378	37	45.1	564	6	AAE38212	Aae38212 Human enz
307	38	46.3	344	4	AAW79272	Aaw79272 Corynebac	379	37	45.1	582	8	ADK24823	Adk24823 plant ful
308	38	46.3	344	4	AAW92789	Aaw92789 C. glutami	380	37	45.1	585	2	AAV13375	Aav13375 Amino aci
309	38	46.3	344	6	ABU26122	Abu26122 Protein e	381	37	45.1	587	8	ADOL3883	Adol3883 P. amagag
310	38	46.3	357	3	AAG47948	Aag47948 Arabidops	382	37	45.1	604	9	AEA60269	Aea60269 Streptoco
311	38	46.3	358	3	AAG20844	Aag20844 Arabidops	383	37	45.1	604	9	AEA60269	Aea60269 Streptoco
312	38	46.3	360	8	ADO62025	Ado62025 Transcrip	384	37	45.1	610	8	ADH35329	Adh35329 ENZM prot
313	38	46.3	364	7	ADH87100	Adh87100 Enterococ	385	37	45.1	610	8	ABM84753	Abm84753 Human dia
314	38	46.3	367	5	ADP65678	Adp65678 Bifidobac	386	37	45.1	612	8	ABM84752	Abm84752 Human dia
315	38	46.3	379	3	AAG47947	Aag47947 Arabidops	387	37	45.1	615	8	ADY07633	Ady07633 Plant ful
	38	46.3	383	3	AAG47946	Aag47946 Arabidops	388	37	45.1	616	8	ABM84751	Abm84751 Human dia

389	37	45.1	624	3	AAG30591	Aeg30591	Arabidops	462	37	45.1	654	6	ABU81251	Human	PRO
390	37	45.1	631	2	AAW77896	Aar77896	Bacterial	463	37	45.1	654	6	ABR60048	Human	sec
391	37	45.1	631	2	AAW08970	Aaw08970	Amino aci	464	37	45.1	654	6	ABR67783	Human	sec
392	37	45.1	631	2	RAY51784	Aay51784	H. influe	465	37	45.1	654	6	ABR65171	Human	sec
393	37	45.1	631	2	AAW54128	Aaw54128	H. influe	466	37	45.1	654	6	ABR68393	Human	sec
394	37	45.1	631	3	AAW80366	Aay80366	H. influe	467	37	45.1	654	6	ABR71805	Human	sec
395	37	45.1	648	8	ADO13885	Ado13885	P. amagas	468	37	45.1	654	6	ABR85285	Human	PRO
396	37	45.1	654	4	ADC78497	Adc78497	Human PRO	469	37	45.1	654	6	ABU88975	Human	sec
397	37	45.1	654	4	AAB80243	Aab80243	Human PRO	470	37	45.1	654	6	ABU83055	Human	sec
398	37	45.1	654	4	RAE06593	Aae06593	Human PRO	471	37	45.1	654	6	ABU94911	Novel hum	
399	37	45.1	654	4	AAU29036	Aau29036	Human PRO	472	37	45.1	654	6	ABU90459	Novel hum	
400	37	45.1	654	4	ABU58412	Abu58412	Human PRO	473	37	45.1	654	6	ABU83970	Human	sec
401	37	45.1	654	6	ABU71621	Abu71621	Human PRO	474	37	45.1	654	6	ABU93621	Novel hum	
402	37	45.1	654	6	ABU87960	Abu87960	Novel hum	475	37	45.1	654	6	ABR64866	Human	sec
403	37	45.1	654	6	ABU84275	Abu84275	Human sec	476	37	45.1	654	6	ABR68698	Human	sec
404	37	45.1	654	6	ABR66149	Abr66149	Human sec	477	37	45.1	654	6	ABO06514	Human	sec
405	37	45.1	654	6	ABR65539	Abr65539	Human sec	478	37	45.1	654	6	ABR99059	Human	sec
406	37	45.1	654	6	ABU99479	Abu99479	Human sec	479	37	45.1	654	6	ABU56943	Human	PRO
407	37	45.1	654	6	ABU82718	Abu82718	Human PRO	480	37	45.1	654	6	ABU64530	Human	sec
408	37	45.1	654	6	ABU89839	Abu89839	Novel hum	481	37	45.1	654	6	ABU85895	Novel hum	
409	37	45.1	654	6	ABU71476	Abu71476	Human PRO	482	37	45.1	654	6	ABU67376	Human	sec
410	37	45.1	654	6	ABR68088	Abr68088	Human sec	483	37	45.1	654	6	ABU82182	Novel hum	
411	37	45.1	654	6	ABU96141	Abu96141	Novel hum	484	37	45.1	654	6	ABU87193	Human	PRO
412	37	45.1	654	6	ABU92572	Abu92572	Human sec	485	37	45.1	654	6	ABU83665	Human	sec
413	37	45.1	654	6	ABO08649	Abo08649	Human sec	486	37	45.1	654	6	ABO08039	Human	PRO
414	37	45.1	654	6	ABO02701	Abo02701	Human sec	487	37	45.1	654	6	ABO14896	Human	sec
415	37	45.1	654	6	ABR74855	Abr74855	Human sec	488	37	45.1	654	6	ABU81750	Novel hum	
416	37	45.1	654	6	ABR94617	Abr94617	Human sec	489	37	45.1	654	6	ABU65914	Novel hum	
417	37	45.1	654	6	ABU85590	Abu85590	Human PRO	490	37	45.1	654	6	ABR59743	Human	sec
418	37	45.1	654	6	ABU98750	Abu98750	Novel hum	491	37	45.1	654	6	ABU93931	Novel hum	
419	37	45.1	654	6	ABU97965	Abu97965	Novel hum	492	37	45.1	654	6	ABU99784	Novel hum	
420	37	45.1	654	6	ABU91671	Abu91671	Novel hum	493	37	45.1	654	6	ABR66454	Human	sec
421	37	45.1	654	6	ABU71922	Abu71922	Human sec	494	37	45.1	654	6	ABR90872	Human	sec
422	37	45.1	654	6	ABU89364	Abu89364	Human PRO	495	37	45.1	654	6	ABU94299	Human	PRO
423	37	45.1	654	6	ABU86205	Abu86205	Human sec	496	37	45.1	654	6	ABU79181	Human	PRO
424	37	45.1	654	6	ABU67418	Abu67418	Human sec	497	37	45.1	654	6	ABU86510	Human	sec
425	37	45.1	654	6	ABU80446	Abu80446	Human PRO	498	37	45.1	654	6	ABU86815	Novel hum	
426	37	45.1	654	6	ABO01805	Abo01805	Novel hum	499	37	45.1	654	6	ABU94604	Human	PRO
427	37	45.1	654	6	ABR99364	Abr99364	Human sec	500	37	45.1	654	6	ABO04531	Human	PRO
428	37	45.1	654	6	ABR98754	Abr98754	Human sec	501	37	45.1	654	6	ABR70280	Human	sec
429	37	45.1	654	6	ABO16277	Abo16277	Human sec	502	37	45.1	654	6	ABU98445	Human	PRO
430	37	45.1	654	6	ABO192177	Abo192177	Human sec	503	37	45.1	654	6	ABR65844	Human	sec
431	37	45.1	654	6	ABO18818	Abo18818	Human sec	504	37	45.1	654	6	ABR64561	Human	sec
432	37	45.1	654	6	ABR78239	Abr78239	Human sec	505	37	45.1	654	6	ABU15632	Protein e	
433	37	45.1	654	6	ABU84975	Abu84975	Novel hum	506	37	45.1	654	6	ABU79486	Human	PRO
434	37	45.1	654	6	ABO00114	Abo00114	Novel hum	507	37	45.1	654	6	ABU92877	Human	sec
435	37	45.1	654	6	ABO11446	Abo11446	Human sec	508	37	45.1	654	6	ABU95836	Human	PRO
436	37	45.1	654	6	ABO02091	Abo02091	Human sec	509	37	45.1	654	6	ABU91056	Novel hum	
437	37	45.1	654	6	ABU54378	Abu54378	Human sec	510	37	45.1	654	6	ABU90149	Novel hum	
438	37	45.1	654	6	ABU88665	Abu88665	Novel hum	511	37	45.1	654	6	ABO09564	Human	sec
439	37	45.1	654	6	ABU83360	Abu83360	Human sec	512	37	45.1	654	6	ABO10836	Human	sec
440	37	45.1	654	6	ABO06161	Abo06161	Novel hum	513	37	45.1	654	6	ABR70890	Human	PRO
441	37	45.1	654	6	ABR66759	Abr66759	Human sec	514	37	45.1	654	6	ABU87498	Human	PRO
442	37	45.1	654	6	ABR59197	Abr59197	Human sec	515	37	45.1	654	6	ABU91366	Human	PRO
443	37	45.1	654	6	ABO09259	Abo09259	Human sec	516	37	45.1	654	6	ABU84580	Human	sec
444	37	45.1	654	6	ABO19123	Abo19123	Novel hum	517	37	45.1	654	6	ABR69670	Human	sec
445	37	45.1	654	6	ABO11141	Abo11141	Human sec	518	37	45.1	654	6	ABU80047	Human	PRO
446	37	45.1	654	6	ABR66759	Abr66759	Human sec	519	37	45.1	654	6	ABU69653	Novel hum	
447	37	45.1	654	6	ABO15972	Abo15972	Human sec	520	37	45.1	654	6	ABU93316	Human	PRO
448	37	45.1	654	6	ABO13678	Abo13678	Human sec	521	37	45.1	654	6	ABO09869	Human	sec
449	37	45.1	654	6	ABO47393	Abo47393	Human sec	522	37	45.1	654	6	ABO08954	Human	sec
450	37	45.1	654	6	ABU65581	Abu65581	Human sec	523	37	45.1	654	6	ABU10522	Human	sec
451	37	45.1	654	6	ABO07429	Abo07429	Human PRO	524	37	45.1	654	6	ABU95531	Human	PRO
452	37	45.1	654	6	ABO03616	Abo03616	Human sec	525	37	45.1	654	6	ABU95531	Novel hum	
453	37	45.1	654	6	ABR67064	Abr67064	Human sec	526	37	45.1	654	6	ABR70585	Human	sec
454	37	45.1	654	6	ABO15667	Abo15667	Human sec	527	37	45.1	654	6	ABO04936	Novel hum	
455	37	45.1	654	6	ABU55948	Abu55948	Human sec	528	37	45.1	654	6	ABO08344	Human	sec
456	37	45.1	654	6	ABU65276	Abu65276	Human PRO	529	37	45.1	654	6	ABO14835	Human	sec
457	37	45.1	654	6	ABU95221	Abu95221	Novel hum	530	37	45.1	654	6	ABO05551	Human	sec
458	37	45.1	654	6	ABU71124	Abu71124	Human PRO	531	37	45.1	654	6	ABR73940	Human	sec
459	37	45.1	654	6	ABO07734	Abo07734	Human PRO	532	37	45.1	654	6	ABR95532	Human	sec
460	37	45.1	654	6	ABR69975	Abr69975	Human sec	533	37	45.1	654	6	ABR80829	Human	sec
461	37	45.1	654	6	ABR69308	Abr69308	Human sec	534	37	45.1	654	6	ABR81134	Human	sec
					ABO01449	Abo01449	Human PRO								

535	37	45.1	654	6	ABM00830	Human sec	Abm00830	Human sec	608	37	45.1	654	6	ABM28093	Human sec
536	37	45.1	654	6	ABR88432	Human sec	Abm88432	Human sec	609	37	45.1	654	6	ABO32092	Human sec
537	37	45.1	654	6	ABM77253	Human sec	Abm77253	Human sec	610	37	45.1	654	6	ABM15219	Human sec
538	37	45.1	654	6	ABO28737	Human sec	Abm28737	Human sec	611	37	45.1	654	6	ABM06374	Human sec
539	37	45.1	654	6	ABO31482	Human sec	Abm31482	Human sec	612	37	45.1	654	6	ABM04185	Human sec
540	37	45.1	654	6	ABM07899	Human sec	Abm07899	Human sec	613	37	45.1	654	6	ABM22298	Human sec
541	37	45.1	654	6	ABO40379	Human sec	Abm40379	Human sec	614	37	45.1	654	6	ABM07594	Human sec
542	37	45.1	654	6	ABO35804	Human PRO	Abm35804	Human PRO	615	37	45.1	654	6	ABO40684	Human sec
543	37	45.1	654	6	ABO43943	Human PRO	Abm43943	Human PRO	616	37	45.1	654	6	ABM35331	Human sec
544	37	45.1	654	6	ADA77778	Human sec	Ada77778	Human sec	617	37	45.1	654	6	ABM33094	Human sec
545	37	45.1	654	6	ABM24738	Human sec	Abm24738	Human sec	618	37	45.1	654	6	ABO52620	Human PRO
546	37	45.1	654	6	ABD29382	Human sec	Abd29382	Human sec	619	37	45.1	654	6	ABO50180	Human sec
547	37	45.1	654	6	ABO03006	Human sec	Abm03006	Human sec	620	37	45.1	654	6	ABU99174	Human sec
548	37	45.1	654	6	ABR90262	Human sec	Abm90262	Human sec	621	37	45.1	654	6	ABO44226	Human sec
549	37	45.1	654	6	ABM17176	Human sec	Abm17176	Human sec	622	37	45.1	654	6	ABO05856	Human sec
550	37	45.1	654	6	ABR94922	Human sec	Abm94922	Human sec	623	37	45.1	654	6	ABO34847	Human PRO
551	37	45.1	654	6	ABR95227	Human sec	Abm95227	Human sec	624	37	45.1	654	6	ABM18396	Human sec
552	37	45.1	654	6	ABO21465	Human sec	Abm21465	Human sec	625	37	45.1	654	6	ADA16213	Human sec
553	37	45.1	654	6	ABR97729	Human sec	Abm97729	Human sec	626	37	45.1	654	6	ABR97424	Human sec
554	37	45.1	654	6	ABR87517	Human sec	Abm87517	Human sec	627	37	45.1	654	6	ABR80524	Human sec
555	37	45.1	654	6	ABM77558	Human sec	Abm77558	Human sec	628	37	45.1	654	6	ABM01135	Human sec
556	37	45.1	654	6	ABM27788	Human sec	Abm27788	Human sec	629	37	45.1	654	6	ABR88737	Human sec
557	37	45.1	654	6	ABM06069	Human sec	Abm06069	Human sec	630	37	45.1	654	6	ABM13389	Human sec
558	37	45.1	654	6	ABM03575	Human sec	Abm03575	Human sec	631	37	45.1	654	6	ABM20773	Human sec
559	37	45.1	654	6	ABM35026	Human sec	Abm35026	Human sec	632	37	45.1	654	6	ABO41904	Human sec
560	37	45.1	654	6	ABM26263	Human sec	Abm26263	Human sec	633	37	45.1	654	6	ABO42514	Human sec
561	37	45.1	654	6	ABO48045	Human sec	Abm48045	Human sec	634	37	45.1	654	6	ABM10034	Human sec
562	37	45.1	654	6	ABR92787	Human sec	Abm92787	Human sec	635	37	45.1	654	6	ABO38549	Human sec
563	37	45.1	654	6	ABO24548	Human sec	Abm24548	Human sec	636	37	45.1	654	6	ABM32789	Human sec
564	37	45.1	654	6	ABM11559	Human sec	Abm11559	Human sec	637	37	45.1	654	6	ABM22603	Human sec
565	37	45.1	654	6	ABM02660	Human sec	Abm02660	Human sec	638	37	45.1	654	6	ABM74814	Human sec
566	37	45.1	654	6	ABM15956	Human sec	Abm15956	Human sec	639	37	45.1	654	6	ADA79570	Human sec
567	37	45.1	654	6	ABO27517	Human sec	Abm27517	Human sec	640	37	45.1	654	6	ABR96204	Human sec
568	37	45.1	654	6	ABM29008	Human sec	Abm29008	Human sec	641	37	45.1	654	6	ABM02355	Human sec
569	37	45.1	654	6	ABM06984	Human sec	Abm06984	Human sec	642	37	45.1	654	6	ABR86297	Human sec
570	37	45.1	654	6	ABM21078	Human sec	Abm21078	Human sec	643	37	45.1	654	6	ABR86602	Human sec
571	37	45.1	654	6	ABM09424	Human sec	Abm09424	Human sec	644	37	45.1	654	6	ABM16566	Human sec
572	37	45.1	654	6	ABO41294	Human sec	Abm41294	Human sec	645	37	45.1	654	6	ABM29618	Human sec
573	37	45.1	654	6	ABO36109	Human PRO	Abm36109	Human PRO	646	37	45.1	654	6	ABO29042	Human sec
574	37	45.1	654	6	ABO43638	Human PRO	Abm43638	Human PRO	647	37	45.1	654	6	ABM23823	Human sec
575	37	45.1	654	6	ABM76338	Human sec	Abm76338	Human sec	648	37	45.1	654	6	ABM23213	Human sec
576	37	45.1	654	6	ABM76034	Human sec	Abm76034	Human sec	649	37	45.1	654	6	ABM21993	Human sec
577	37	45.1	654	6	ABM25653	Human sec	Abm25653	Human sec	650	37	45.1	654	6	ABO37634	Human sec
578	37	45.1	654	6	ABM25958	Human sec	Abm25958	Human sec	651	37	45.1	654	6	ABM28398	Human sec
579	37	45.1	654	6	ABO03311	Human sec	Abm03311	Human sec	652	37	45.1	654	6	ABM28703	Human sec
580	37	45.1	654	6	ABO02396	Human sec	Abm02396	Human sec	653	37	45.1	654	6	ABM66347	Human sec
581	37	45.1	654	6	ABR90567	Human sec	Abm90567	Human sec	654	37	45.1	654	6	ABM66347	Human sec
582	37	45.1	654	6	ABR73635	Human sec	Abm73635	Human sec	655	37	45.1	654	6	ABM75729	Human sec
583	37	45.1	654	6	ABO16887	Human sec	Abm16887	Human sec	656	37	45.1	654	6	ABM34009	Human sec
584	37	45.1	654	6	ABR94312	Human sec	Abm94312	Human sec	657	37	45.1	654	6	ABM34314	Human sec
585	37	45.1	654	6	ABR75819	Human sec	Abm75819	Human sec	658	37	45.1	654	6	ABO20245	Human sec
586	37	45.1	654	6	ADA18238	Human sec	Ada18238	Human sec	659	37	45.1	654	6	ABO21160	Human sec
587	37	45.1	654	6	ABO32787	Human sec	Abm32787	Human sec	660	37	45.1	654	6	ABO22075	Human sec
588	37	45.1	654	6	ABR71195	Human sec	Abm71195	Human sec	661	37	45.1	654	6	ABR96509	Human sec
589	37	45.1	654	6	ABR93092	Human sec	Abm93092	Human sec	662	37	45.1	654	6	ABR85687	Human sec
590	37	45.1	654	6	ABR93397	Human sec	Abm93397	Human sec	663	37	45.1	654	6	ABR99669	Human sec
591	37	45.1	654	6	ABR87822	Human sec	Abm87822	Human sec	664	37	45.1	654	6	ABM00525	Human sec
592	37	45.1	654	6	ABO27822	Human sec	Abm27822	Human sec	665	37	45.1	654	6	ABO29652	Human sec
593	37	45.1	654	6	ABO29957	Human sec	Abm29957	Human sec	666	37	45.1	654	6	ABM23518	Human sec
594	37	45.1	654	6	ABO33166	Human PRO	Abm33166	Human PRO	667	37	45.1	654	6	ABM29313	Human sec
595	37	45.1	654	6	ABM04854	Human sec	Abm04854	Human sec	668	37	45.1	654	6	ABO38244	Human sec
596	37	45.1	654	6	ABM08814	Human sec	Abm08814	Human sec	669	37	45.1	654	6	ABO45544	Human PRO
597	37	45.1	654	6	ABO36414	Human sec	Abm36414	Human sec	670	37	45.1	654	6	ABM20468	Human sec
598	37	45.1	654	6	ABO35499	Human PRO	Abm35499	Human PRO	671	37	45.1	654	6	ADA42358	Human sec
599	37	45.1	654	6	ABO39464	Human sec	Abm39464	Human sec	672	37	45.1	654	6	ADA81297	Human sec
600	37	45.1	654	6	ABM10339	Human sec	Abm10339	Human sec	673	37	45.1	654	6	ABO16582	Human sec
601	37	45.1	654	6	ABM11864	Human sec	Abm11864	Human sec	674	37	45.1	654	6	ABO18208	Human PRO
602	37	45.1	654	6	ABO52010	Human PRO	Abm52010	Human PRO	675	37	45.1	654	6	ABO22635	Human PRO
603	37	45.1	654	6	ABO52315	Human PRO	Abm52315	Human PRO	676	37	45.1	654	6	ABO22940	Human PRO
604	37	45.1	654	6	ABO23633	Human sec	Abm23633	Human sec	677	37	45.1	654	6	ABR92482	Human sec
605	37	45.1	654	6	ABR97119	Human sec	Abm97119	Human sec	678	37	45.1	654	6	ABR81439	Human sec
606	37	45.1	654	6	ABR86907	Human sec	Abm86907	Human sec	679	37	45.1	654	6	ABO17525	Human PRO
607	37	45.1	654	6	ABM10949	Human sec	Abm10949	Human sec	680	37	45.1	654	6	ABM77863	Human PRO

681	37	45.1	654	6	ABR99652	ABr89652 Human sec	754	37	45.1	654	7	ABO50790	ABo50790 Human sec
682	37	45.1	654	6	ABR26568	ABm26568 Human sec	755	37	45.1	654	7	ABO05246	ABo05246 Human sec
683	37	45.1	654	6	ABM13694	ABm13694 Human sec	756	37	45.1	654	7	ABR74550	ABr74550 Human sec
684	37	45.1	654	6	ABO28432	ABo28432 Human sec	757	37	45.1	654	7	ABR77029	ABr77029 Human sec
685	37	45.1	654	6	ABO30262	ABo30262 Human sec	758	37	45.1	654	7	ADA16637	Ada16637 Human sec
686	37	45.1	654	6	ABM07289	ABm07289 Human sec	759	37	45.1	654	7	ABM17786	ABm17786 Human sec
687	37	45.1	654	6	ABM03880	ABm03880 Human sec	760	37	45.1	654	7	ABR95837	ABr95837 Human sec
688	37	45.1	654	6	ABO37024	ABo37024 Human sec	761	37	45.1	654	7	ADA13066	Ada13066 Human sec
689	37	45.1	654	6	ABO41599	ABo41599 Human sec	762	37	45.1	654	7	ABO21770	ABo21770 Human sec
690	37	45.1	654	6	ABO35194	ABo35194 Human PRO	763	37	45.1	654	7	ABO119940	ABo119940 Human sec
691	37	45.1	654	6	ABM25043	ABm25043 Human sec	764	37	45.1	654	7	ABO24243	ABo24243 Human sec
692	37	45.1	654	6	ABO47435	ABo47435 Human sec	765	37	45.1	654	7	ABR85992	ABr85992 Human sec
693	37	45.1	654	6	ABO47740	ABo47740 Human sec	766	37	45.1	654	7	ABM10644	ABm10644 Human sec
694	37	45.1	654	6	ABO48350	ABo48350 Human sec	767	37	45.1	654	7	ABM76643	ABm76643 Human sec
695	37	45.1	654	6	ABO51400	ABo51400 Human PRO	768	37	45.1	654	7	ABR89347	ABr89347 Human sec
696	37	45.1	654	6	ABO51705	ABo51705 Human PRO	769	37	45.1	654	7	ABM12474	ABm12474 Human sec
697	37	45.1	654	6	ABO50485	ABo50485 Human sec	770	37	45.1	654	7	ABM05764	ABm05764 Human sec
698	37	45.1	654	6	ABR79609	ABr79609 Human sec	771	37	45.1	654	7	ABO34889	ABo34889 Human PRO
699	37	45.1	654	6	ABM16871	ABm16871 Human sec	772	37	45.1	654	7	ABM02965	ABm02965 Human sec
700	37	45.1	654	6	ABO17903	ABo17903 Human sec	773	37	45.1	654	7	ABM18943	ABm18943 Human sec
701	37	45.1	654	6	ABO20855	ABo20855 Human sec	774	37	45.1	654	7	ABM19248	ABm19248 Human sec
702	37	45.1	654	6	ABR96814	ABr96814 Human sec	775	37	45.1	654	7	ABO46459	ABo46459 Human PRO
703	37	45.1	654	6	ABM12169	ABm12169 Human sec	776	37	45.1	654	7	ABO48960	ABo48960 Human sec
704	37	45.1	654	6	ABM16261	ABm16261 Human sec	777	37	45.1	654	7	ADA41934	Ada41934 Human sec
705	37	45.1	654	6	ABM24128	ABm24128 Human sec	778	37	45.1	654	7	ABR69003	ABr69003 Human sec
706	37	45.1	654	6	ABM14609	ABm14609 Human sec	779	37	45.1	654	7	ABR89042	ABr89042 Human sec
707	37	45.1	654	6	ABM04490	ABm04490 Human sec	780	37	45.1	654	7	ABR72415	ABr72415 Human sec
708	37	45.1	654	6	ABM06679	ABm06679 Human sec	781	37	45.1	654	7	ABR74245	ABr74245 Human sec
709	37	45.1	654	6	ABM09119	ABm09119 Human sec	782	37	45.1	654	7	ABO18513	ABo18513 Human sec
710	37	45.1	654	6	ABO39159	ABo39159 Human sec	783	37	45.1	654	7	ADA17281	Ada17281 Human sec
711	37	45.1	654	6	ABM75424	ABm75424 Human sec	784	37	45.1	654	7	ABR80219	ABr80219 Human sec
712	37	45.1	654	6	ABM25348	ABm25348 Human sec	785	37	45.1	654	7	ABM01440	ABm01440 Human sec
713	37	45.1	654	6	ABM19858	ABm19858 Human sec	786	37	45.1	654	7	ABM02050	ABm02050 Human sec
714	37	45.1	654	6	ABO46764	ABo46764 Human PRO	787	37	45.1	654	7	ABR87212	ABr87212 Human sec
715	37	45.1	654	6	ABO47069	ABo47069 Human PRO	788	37	45.1	654	7	ABM12779	ABm12779 Human sec
716	37	45.1	654	6	ADA83095	Ada83095 Human sec	789	37	45.1	654	7	ABM30533	ABm30533 Human sec
717	37	45.1	654	6	ABR71500	ABr71500 Human sec	790	37	45.1	654	7	ABM24433	ABm24433 Human sec
718	37	45.1	654	6	ABR72110	ABr72110 Human sec	791	37	45.1	654	7	ABO29347	ABo29347 Human sec
719	37	45.1	654	6	ABR98449	ABr98449 Human sec	792	37	45.1	654	7	ABO31177	ABo31177 Human sec
720	37	45.1	654	6	ABO06819	ABo06819 Human sec	793	37	45.1	654	7	ABM14304	ABm14304 Human sec
721	37	45.1	654	6	ABR84772	ABr84772 Human sec	794	37	45.1	654	7	ABM09729	ABm09729 Human sec
722	37	45.1	654	6	ABR73330	ABr73330 Human sec	795	37	45.1	654	7	ABO38854	ABo38854 Human sec
723	37	45.1	654	6	ABR76424	ABr76424 Human sec	796	37	45.1	654	7	ABM34619	ABm34619 Human sec
724	37	45.1	654	6	ABR73025	ABr73025 Human sec	797	37	45.1	654	7	ABO51095	ABo51095 Human sec
725	37	45.1	654	6	ABM18091	ABm18091 Human sec	798	37	45.1	654	7	ADA42784	Ada42784 Human sec
726	37	45.1	654	6	ABO20550	ABo20550 Human sec	799	37	45.1	654	7	ABO03921	ABo03921 Human sec
727	37	45.1	654	6	ABO25293	ABo25293 Human PRO	800	37	45.1	654	7	ABO10391	ABo10391 Human PRO
728	37	45.1	654	6	ABO25598	ABo25598 Human PRO	801	37	45.1	654	7	ABR77634	ABr77634 Human sec
729	37	45.1	654	6	ABR94007	ABr94007 Human sec	802	37	45.1	654	7	ABR78844	ABr78844 Human sec
730	37	45.1	654	6	ABR79914	ABr79914 Human sec	803	37	45.1	654	7	ABO23938	ABo23938 Human sec
731	37	45.1	654	6	ABM11254	ABm11254 Human sec	804	37	45.1	654	7	ABR93702	ABr93702 Human sec
732	37	45.1	654	6	ABO32861	ABo32861 Human sec	805	37	45.1	654	7	ABM01745	ABm01745 Human sec
733	37	45.1	654	6	ABO30567	ABo30567 Human sec	806	37	45.1	654	7	ABM78168	ABm78168 Human sec
734	37	45.1	654	6	ABO30872	ABo30872 Human sec	807	37	45.1	654	7	ABR89957	ABr89957 Human sec
735	37	45.1	654	6	ABM27178	ABm27178 Human sec	808	37	45.1	654	7	ABM27483	ABm27483 Human sec
736	37	45.1	654	6	ABM29923	ABm29923 Human sec	809	37	45.1	654	7	ABM13084	ABm13084 Human sec
737	37	45.1	654	6	ABM05459	ABm05459 Human sec	810	37	45.1	654	7	ABO31787	ABo31787 Human sec
738	37	45.1	654	6	ABM15524	ABm15524 Human sec	811	37	45.1	654	7	ABM13999	ABm13999 Human sec
739	37	45.1	654	6	ABM08509	ABm08509 Human sec	812	37	45.1	654	7	ABM08204	ABm08204 Human sec
740	37	45.1	654	6	ABO42209	ABo42209 Human sec	813	37	45.1	654	7	ABO40074	ABo40074 Human sec
741	37	45.1	654	6	ABO37939	ABo37939 Human sec	814	37	45.1	654	7	ABM74509	ABm74509 Human sec
742	37	45.1	654	6	ABO45849	ABo45849 Human PRO	815	37	45.1	654	7	ABM33704	ABm33704 Human sec
743	37	45.1	654	6	ABM66652	ABm66652 Human sec	816	37	45.1	654	7	ABM20163	ABm20163 Human sec
744	37	45.1	654	6	ADB20138	ABd20138 Human sec	817	37	45.1	654	7	ABO48655	ABo48655 Human sec
745	37	45.1	654	6	ABM19553	ABm19553 Human sec	818	37	45.1	654	7	ABR72720	ABr72720 Human sec
746	37	45.1	654	6	ABO49265	ABo49265 Human sec	819	37	45.1	654	7	ABO15362	ABo15362 Human sec
747	37	45.1	654	6	ABO49570	ABo49570 Human sec	820	37	45.1	654	7	ABR85077	ABr85077 Human sec
748	37	45.1	654	6	ADA78390	Ada78390 Human sec	821	37	45.1	654	7	ABO15057	ABo15057 Human sec
749	37	45.1	654	6	ABR88127	ABr88127 Human sec	822	37	45.1	654	7	ABO17192	ABo17192 Human sec
750	37	45.1	654	6	ABM26873	ABm26873 Human sec	823	37	45.1	654	7	ABm17481	ABm17481 Human sec
751	37	45.1	654	6	ABM03270	ABm03270 Human sec	824	37	45.1	654	7	ABR85382	ABr85382 Human sec
752	37	45.1	654	6	ABO39769	ABo39769 Human sec	825	37	45.1	654	7	ABO17586	ABo17586 Human PRO
753	37	45.1	654	7	ABO49875	ABo49875 Human sec	826	37	45.1	654	7	ABM76948	ABm76948 Human sec

827	37	45.1	654	7	ABO28127	Human sec	900	37	45.1	654	8	ADI18659	Human sec
828	37	45.1	654	7	ABM22908	Human sec	901	37	45.1	654	8	ADI165379	Human sec
829	37	45.1	654	7	ABM30228	Human sec	902	37	45.1	654	8	ADI37642	Human sec
830	37	45.1	654	7	ABM21688	Human sec	903	37	45.1	654	8	ADH97438	Human sec
831	37	45.1	654	7	ABM21383	Human sec	904	37	45.1	654	8	ADH165806	Human sec
832	37	45.1	654	7	ABM14914	Human sec	905	37	45.1	654	8	ADH60549	Human sec
833	37	45.1	654	7	ABO40989	Human sec	906	37	45.1	654	8	ADH99606	Human sec
834	37	45.1	654	7	ABO36719	Human sec	907	37	45.1	654	8	ADL08799	Human sec
835	37	45.1	654	7	ABO37329	Human sec	908	37	45.1	654	8	ADJ54664	Human sec
836	37	45.1	654	7	ABM75119	Human sec	909	37	45.1	654	8	ADM25140	Human sec
837	37	45.1	654	7	ABM33399	Human sec	910	37	45.1	654	8	ADM29890	Human sec
838	37	45.1	654	7	ABO46154	Human PRO	911	37	45.1	654	8	ADJ64435	Human PRO
839	37	45.1	654	7	ADA82461	Human sec	912	37	45.1	654	8	ADK47677	Streptoco
840	37	45.1	654	7	ABM31753	Human sec	913	37	45.1	654	8	ADM31331	Novel hum
841	37	45.1	654	7	ABM31143	Human sec	914	37	45.1	654	8	ADM36378	Novel hum
842	37	45.1	654	7	ADB77703	Human sec	915	37	45.1	654	8	ADM40183	Novel hum
843	37	45.1	654	7	ADB74839	Human sec	916	37	45.1	654	8	ADO06212	Human PRO
844	37	45.1	654	7	ADB85769	Human sec	917	37	45.1	654	8	ADM37791	Novel hum
845	37	45.1	654	7	ABM32058	Human sec	918	37	45.1	654	8	ADR11064	Human sec
846	37	45.1	654	7	ABM32363	Human sec	919	37	45.1	654	8	ADR17973	Human sec
847	37	45.1	654	7	ABM31448	Human sec	920	37	45.1	654	8	ADH74612	Human sec
848	37	45.1	654	7	ABM30838	Human sec	921	37	45.1	654	8	ADT03649	Human sec
849	37	45.1	654	7	ADC28485	Human sec	922	37	45.1	654	9	AEA37887	Human sec
850	37	45.1	654	7	ADC39685	Human sec	923	37	45.1	654	9	AED23921	Human sec
851	37	45.1	654	7	ADC40199	Human sec	924	37	45.1	654	9	AED50021	Novel hum
852	37	45.1	654	7	ADC19023	Human sec	925	37	45.1	654	10	ABE69090	Beta-gala
853	37	45.1	654	7	ADC34323	Human sec	926	37	45.1	681	7	ABO80085	Pseudomon
854	37	45.1	654	7	ADC29378	Human sec	927	37	45.1	683	8	ABM84750	Human dia
855	37	45.1	654	7	ADC28909	Human sec	928	37	45.1	722	8	ADO13886	P. amagas
856	37	45.1	654	7	ADC40794	Human sec	929	37	45.1	761	4	ABG02677	Novel hum
857	37	45.1	654	7	ADC19451	Human sec	930	37	45.1	764	4	ABG09702	Novel hum
858	37	45.1	654	7	ADC33899	Human sec	931	37	45.1	780	3	ABG30590	Arabidops
859	37	45.1	654	7	ADC12969	Human sec	932	37	45.1	780	7	ADB95076	A. thalia
860	37	45.1	654	7	ADC12421	Human sec	933	37	45.1	782	4	ABG16643	Novel hum
861	37	45.1	654	7	ADD05499	Human sec	934	37	45.1	786	6	ABU02588	S. pneumo
862	37	45.1	654	7	ADD04976	Human sec	935	37	45.1	803	4	AAU34101	Staphyloc
863	37	45.1	654	7	ADD03982	Human sec	936	37	45.1	804	2	AAW22709	Leucyl-tr
864	37	45.1	654	7	ADD03558	Human sec	937	37	45.1	804	6	ABU15896	Protein e
865	37	45.1	654	7	ADH34810	Human sec	938	37	45.1	804	6	AAE36398	Staphyloc
866	37	45.1	654	7	ADG02494	Novel hum	939	37	45.1	804	9	ADW94769	prolifera
867	37	45.1	654	7	ADG01201	Novel hum	940	37	45.1	805	4	AAU36693	Staphyloc
868	37	45.1	654	7	ADF95376	Novel hum	941	37	45.1	805	6	ADA89537	Staphyloc
869	37	45.1	654	7	ADG12191	Novel hum	942	37	45.1	805	6	ABM71188	Staphyloc
870	37	45.1	654	7	ADH08851	Human PRO	943	37	45.1	834	8	ADSL6669	B. hensel
871	37	45.1	654	7	ADH59293	Human sec	944	37	45.1	843	8	ADM92249	S. pneumon
872	37	45.1	654	7	ADH138072	Human sec	945	37	45.1	844	4	AAU37922	Streptoco
873	37	45.1	654	7	ADJ26340	Human sec	946	37	45.1	857	6	ABU46277	Protein e
874	37	45.1	654	7	ADL32632	Novel hum	947	37	45.1	1059	8	ADK99177	Streptoco
875	37	45.1	654	7	ADM30166	Novel hum	948	37	45.1	1205	9	ABM97591	M. xanthu
876	37	45.1	654	8	ADE79255	Human sec	949	37	45.1	1597	4	ABB60567	Drosophill
877	37	45.1	654	8	ADE79679	Human sec	950	36.5	44.5	72	8	ABO55120	Human gen
878	37	45.1	654	8	ADE79679	Human sec	951	36.5	44.5	87	4	AAO10694	Human pol
879	37	45.1	654	8	ADE73355	Human sec	952	36.5	44.5	116	4	ABG00431	Novel hum
880	37	45.1	654	8	ADE73355	Human sec	953	36.5	44.5	126	8	ADQ36997	Cell prol
881	37	45.1	654	8	ADE74775	Human sec	954	36.5	44.5	126	8	ADQ15667	Rice stre
882	37	45.1	654	8	ADE99444	Human sec	955	36.5	44.5	443	8	ADSL25494	Bacterial
883	37	45.1	654	8	ADE98563	Human sec	956	36.5	44.5	443	8	ADS25995	Bacterial
884	37	45.1	654	8	ADE98990	Human sec	957	36.5	44.5	443	8	ADS22580	Bacterial
885	37	45.1	654	8	ADG40460	Human sec	958	36.5	44.5	581	4	AAU41714	Propionib
886	37	45.1	654	8	ADF73854	Human sec	959	36.5	44.5	581	6	ABM38233	Propionib
887	37	45.1	654	8	ADF95988	Novel hum	960	36	43.9	9	9	ADU70657	Human hep
888	37	45.1	654	8	ADG73430	Novel hum	961	36	43.9	35	6	ABU01244	S. pneumo
889	37	45.1	654	8	ADG04259	Human sec	962	36	43.9	49	4	ABG01573	Novel hum
890	37	45.1	654	8	ADG00419	Novel hum	963	36	43.9	65	4	AAU65541	Propionib
891	37	45.1	654	8	ADG82675	Human PRO	964	36	43.9	65	6	ABM62060	Propionib
892	37	45.1	654	8	ADG92273	Human sec	965	36	43.9	66	4	AAO03574	Human pol
893	37	45.1	654	8	ADG92700	Human sec	966	36	43.9	75	3	AAAG21635	Arabidops
894	37	45.1	654	8	ADH25956	Novel hum	967	36	43.9	90	4	AAO06864	Human pol
895	37	45.1	654	8	ADH32925	Human PRO	968	36	43.9	106	4	ABG26789	Novel hum
896	37	45.1	654	8	ADH20489	Human sec	969	36	43.9	112	4	AAU07425	Human hep
897	37	45.1	654	8	ADH07344	Human sec	970	36	43.9	126	8	ADX89866	Plant ful
898	37	45.1	654	8	ADH59889	Human sec	971	36	43.9	129	4	AAH82623	S. epider
899	37	45.1	654	8	ADH06917	Human sec	972	36	43.9	135	4	AAO05901	Human pol

973 36 43.9 147 4 ABG15872 Novel hum
 974 36 43.9 152 4 AAB45807 Zebratish
 975 36 43.9 152 9 AED08573 Zebratish
 976 36 43.9 153 4 ABG18652 Novel hum
 977 36 43.9 193 4 AAM14443 Peptide #
 978 36 43.9 193 4 AAM15868 Peptide #
 979 36 43.9 193 4 ABB36188 Peptide #
 980 36 43.9 193 4 ABB33391 Peptide #
 981 36 43.9 193 4 AAM28376 Peptide #
 982 36 43.9 193 4 AAM26856 Peptide #
 983 36 43.9 193 4 AAM29680 Peptide #
 984 36 43.9 193 4 ABB28216 Peptide #
 985 36 43.9 193 4 ABB29682 Peptide #
 986 36 43.9 193 4 ABB18850 Protein #
 987 36 43.9 193 4 AAM69353 Human bon
 988 36 43.9 193 4 AAM66570 Human bon
 989 36 43.9 193 4 AAM55668 Human bra
 990 36 43.9 193 4 AAM54176 Human bra
 991 36 43.9 193 4 AAM56966 Human bra
 992 36 43.9 193 4 ABG48238 Human liv
 993 36 43.9 193 4 ABG51029 Human liv
 994 36 43.9 193 4 AAM02170 Peptide #
 995 36 43.9 193 4 AAM04882 Peptide #
 996 36 43.9 193 4 ABB36222 Human pep
 997 36 43.9 193 5 ABG38972 Human pep
 998 36 43.9 209 7 ADC00990 Enterohae
 999 36 43.9 213 6 ABU33837 Protein e
 1000 36 43.9 228 6 ABU19393 Protein e

ALIGNMENTS

RESULT 1
 ADR88214
 ID ADR88214 standard; peptide; 15 AA.

XX ADR88214;

DT 18-NOV-2004 (first entry)

XX Human heparanase epitope pep8.

XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
 KW human; heparanase; enzyme; epitope.

XX Homo sapiens.

OS US2004170631-A1.

XX 02-SEP-2004.

XX 28-NOV-2003; 2003US-00722502.

XX 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

PR 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (VGEL/) VAN GELDER J M.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Herskovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-625084/60.

DR Targeted drug delivery to a heparanase-expressing tissue of a patient,
 XX useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.

Claim 7; SEQ ID NO 8; 58pp; English.

The invention relates to a method of targeted drug delivery to a tissue
 of a patient, the tissue expressing heparanase. The method comprises
 providing a complex of a drug directly or indirectly linked to an anti-
 heparanase antibody, and administering the complex to the patient. In the
 targeted drug delivery, the antibody comprises an antibody or its portion
 capable of specifically binding to at least one epitope of a heparanase
 protein. The composition and methods of the invention are useful for
 diagnosing, preventing or treating conditions associated with heparanase
 catalytic activity (e.g. an inflammatory disorder, wound, scar,
 vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 proliferation, invasion of circulating tumour cells and metastatic
 disease), for purifying heparanase, or for developing drugs for those
 heparanase-associated conditions. The vasculopathy is atherosclerosis,
 restenosis or aneurysm. The cancerous condition is a solid cancer or a
 non-solid cancer. The non-solid cancer is a haematopoietic malignant
 selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 multiple myeloma. The solid cancer is selected from tumours in lip and
 oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 sarcoma. The present sequence is human heparanase epitope.

XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 8; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.4e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SWELGNPNPSFLKKA 15

Db 1 SWELGNPNPSFLKKA 15

RESULT 2

ADT78181

ID ADT78181 standard; peptide; 15 AA.

XX ADT78181;

XX 13-JAN-2005 (first entry)

XX Functional peptide epitope of human heparanase, pep8.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;

KW cell proliferation; cancerous condition; tumour cell invasion;

KW metastatic disease; heparanase-related disorder; inflammatory disorder;

KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.

OS Homo sapiens.

XX US2004213789-A1.

XX PN 28-OCT-2004.

XX PD 22-AUG-2003; 2003US-00645659.

XX PF 02-SEP-1997; 97US-00922170.

XX PR 01-MAY-1998; 98US-00071739.

XX PR 04-NOV-1998; 98US-00186200.

XX PR 19-FEB-2003; 2003US-00368044.

XX (YACOBY) YACOB-ZEEVI O.

XX PA (PERE) PERETZ T.

XX PA (MIRO) MIRON D.

XX PA (SHLO) SHLOMI Y.

XX PA (PECK) PECKER I.

XX PA (AYAL) AYAL-HERSHKOVITZ M.

XX PA (FEIN) FEINSTEIN E.

XX PA (GELD) GELDER J M V.

XX PA (VLOD) VLODAVSKY I.

XX PA (FRIE) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 67; SEQ ID NO 8; 69pp; English.

XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents a functional peptide
 CC epitope of human heparanase.

XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 8; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

Db 1 SWELGNEPNSFLKKA 15

RESULT 3
 ADU70829

XX ID ADU70829 standard; peptide; 15 AA.

XX AC ADU70829;

XX DT 10-FEB-2005 (first entry)

XX DE Human heparanase peptide SEQ ID NO:514.

XX enzyme; heparinase; vaccine; human leukocyte antigen; HLA;

XX immunostimulant; cytostatic; immune disorder; metastasis.

XX OS Homo sapiens.

XX PN EP1479764-A1.

XX PD 24-NOV-2004.

XX PF 19-MAY-2003; 2003EP-00011038.

XX PR 19-MAY-2003; 2003EP-00011038.

XX PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.

XX PI Schirmacher V, Beckhove P, Sommerfeldt N;

XX DR WPI; 2005-014847/02.

XX New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
 PT molecule or its functional derivative, useful for preparing a medicament
 PT for inducing an immune response or for treating metastatic tumors.

XX Disclosure; SEQ ID NO 514; 269pp; English.

XX The invention relates to a novel heparanase peptide that binds to a human
 CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
 CC its functional derivative. A peptide of the invention has immunostimulant
 CC and cytostatic activity, and is used in a vaccine. The heparinase peptide
 CC is useful for preparing a medicament which induces an immune response or
 CC for treating metastatic tumors. The present sequence represents a
 CC heparinase peptide of the invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 9; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

Db 1 SWELGNEPNSFLKKA 15

RESULT 4

AEA42430

XX ID AEA42430 standard; peptide; 15 AA.

XX AC AEA42430;

XX DT 28-JUL-2005 (first entry)

XX DE Human heparanase epitope peptide SEQ ID NO:8.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

XX OS Homo sapiens.

XX AU2004201462-A1.
 XX 06-MAY-2004.
 XX 08-APR-2004; 2004AU-00201462.
 XX 08-APR-2004; 2004AU-00201462.
 XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-HersHKovitz M, Ben-Artzi H;
 PI Feinstein E;
 XX WPI; 2005-173343/19.
 XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.
 XX Claim 7; SEQ ID NO 8; 260pp; English.
 XX The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic
 CC and renal cell carcinoma. The present sequence represents a human
 CC heparanase epitope peptide, which is used in the exemplification of the
 CC present invention.
 XX Sequence 15 AA;
 XX Query Match 100.0%; Score 82; DB 9; Length 15;
 XX Best Local Similarity 100.0%; Pred. No. 2.4e-07;
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNEPNFLKKA 15
 DB 1 SWELGNEPNFLKKA 15
 RESULT 5
 ADR88207
 ID ADR88207 standard; protein; 386 AA.
 XX ADR88207;
 AC ADR88207;
 XX 18-NOV-2004 (first entry)
 XX Human mature heparanase 45 kDa major subunit.
 DE
 XX Targeted drug delivery ; inflammatory disorder; wound; scar;
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;

KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.
 XX Homo sapiens.
 XX US2004170631-A1.
 XX 02-SEP-2004.
 XX 28-NOV-2003; 2003US-00722502.
 XX 02-SEP-1997; 97US-00922170.
 XX 01-MAY-1998; 98US-00071739.
 XX 04-NOV-1998; 98US-00186200.
 XX 19-FEB-2003; 2003US-00368044.
 XX 22-AUG-2003; 2003US-00645659.
 XX (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (VGEL/) VAN GELDER J M.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-HersHKovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 PI Friedmann Y;
 XX WPI; 2004-625084/60.
 XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.
 XX Claim 2; SEQ ID NO 1; 58pp; English.
 XX The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature

CC heparanase.
 XX Sequence 386 AA;
 SQ

Query Match 100.0%; Score 82; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
 Db 62 SWELGNEPNSFLKKA 76

RESULT 6
 ADT78174
 ID ADT78174 standard; protein; 386 AA.
 XX
 AC ADT78174;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE 45kDa subunit of mature processed human heparanase dimer.
 XX
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cystostatic; antiinflammatory; vulnerrary; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
 XX
 OS Homo sapiens.
 XX
 PN US2004213789-A1.
 XX
 PD 28-OCT-2004.
 XX
 PF 22-AUG-2003; 2003US-00645659.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (GELD/) GELDER J M V.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX
 PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Herskovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;
 XX
 XX WPI; 2004-774790/76.
 DR
 XX
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammation or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX
 PS Claim 5; SEQ ID NO 1; 68pp; English.
 XX
 XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal

CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of
 CC mature processed human heparanase dimer.
 XX
 SQ Sequence 386 AA;
 Query Match 100.0%; Score 82; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
 Db 62 SWELGNEPNSFLKKA 76

RESULT 7
 ADY27057
 ID ADY27057 standard; protein; 386 AA.
 XX
 AC ADY27057;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 DE Heparanase inhibitor protein #1.
 XX
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cystostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW heparanase modulator; enzyme purification.
 XX
 OS Homo sapiens.
 XX
 PN WO2005016227-A2.
 XX
 PD 24-FEB-2005.
 XX
 PF 12-AUG-2004; 2004WO-IL000744.
 XX
 PR 14-AUG-2003; 2003US-0494800P.
 PR 12-JAN-2004; 2004US-0535492P.
 XX
 PA (INST-) INSIGHT BIOPHARMACEUTICALS LTD.
 XX
 PI Van-Gelder JM, Miron D;
 XX
 XX WPI; 2005-182203/19.
 DR
 XX
 XX Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.
 XX
 PS Claim 55; SEQ ID NO 33; 211pp; English.
 XX
 XX The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a

CC pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence CC represents a heparanase inhibitor protein used in the scope of the invention.

XX

SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
| | | | | | | | | | | | | | |
Db 62 SWELGNEPNSFLKKA 76

RESULT 8

ID ADZ18995 standard; protein; 386 AA.

XX

AC ADZ18995;

XX

DT 16-JUN-2005 (first entry)

XX

DE Human heparanase consensus cleavage site #2.

XX

KW Enzyme engineering; heparanase; metastasis; autoimmune disease; inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic; immunosuppressive; enzyme.

KW

XX Homo sapiens.

OS

XX WO2005030962-A1.

PN

XX 07-APR-2005.

PD

XX

PF 17-SEP-2004; 2004WO-EP010517.

XX

PR 26-SEP-2003; 2003US-0506479P.

XX

PR 20-JAN-2004; 2004US-0537729P.

XX

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

PA

XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;

PI

XX WPI; 2005-273382/28.

DR

XX

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a nucleotide sequence that encodes mammalian heparanase protein and has two consensus cleavage sites located between specific nucleotide encoding residues.

PT

PT

PT

PT

XX

XX Disclosure; SEQ ID NO 16; 65pp; English.

PS

XX

CC The invention relates to a synthetic nucleic acid molecule that encodes mammalian heparanase protein, where the nucleic acid comprises two consensus cleavage sites recognized by endoproteinase. The sequences are useful for expressing mammalian heparanase in non-mammalian cells and in inhibitor screening assays for the development of therapeutics or pharmaceuticals for inhibiting or treating metastasis, autoimmune disease and/or inflammation. This sequence represents a human heparanase consensus cleavage site used in the scope of the invention.

XX

SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
| | | | | | | | | | | | | | |
Db 62 SWELGNEPNSFLKKA 76

RESULT 9

AEA42423

ID AEA42423 standard; protein; 386 AA.

XX

AC AEA42423;

XX

DT 28-JUL-2005 (first entry)

XX

DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.

XX

KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive; antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic; inflammation; wound healing; scarring; vasculopathy; autoimmune disease; angiogenesis disorder; cancer; tumor; metastasis.

KW

XX Homo sapiens.

OS

XX AU2004201462-A1.

PN

XX 06-MAY-2004.

PD

XX

PF 08-APR-2004; 2004AU-00201462.

XX

XX 08-APR-2004; 2004AU-00201462.

PR

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX

PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;

PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;

PI Feinstein E;

XX

DR WPI; 2005-173343/19.

XX

XX

PT Novel isolated antibody capable of specifically binding to epitope of heparanase protein, useful for preventing and treating heparanase-related disorder such as inflammatory disorder, scars, autoimmune conditions or angiogenesis.

PT

PT

XX

PS Claim 2; SEQ ID NO 1; 260pp; English.

XX

CC The invention relates to an isolated antibody or its portion (I) capable of specifically binding to an epitope of a heparanase protein. Also described: (1) a cell line (II) for producing a monoclonal antibody or its portion, comprising a cell line for producing (I); (2) a pharmaceutical composition comprising (I) and a carrier; and (3) an affinity medium (III) for binding human heparanase polypeptides, comprising (I) immobilized to a chemically inert, insoluble carrier. (I) useful for treating a subject suffering from a pathological condition, which involves administering (I) to the subject. (I) is useful for preventing and treating heparanase-related disorder or condition chosen from inflammatory disorder, wound, scar, vasculopathy, autoimmune condition, angiogenesis, cell proliferation, cancerous condition, tumor cell proliferation, invasion of circulating tumor cells and metastatic disease. (I) is useful for detecting the presence of heparanase polypeptide in a sample. (I) is useful for detecting heparanase-related disease or condition in a subject such as vertebrate, preferably mammal e.g., human. The heparanase-related disorder or condition further includes renal disease or disorder chosen from diabetic nephropathy, glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome and renal cell carcinoma. The present sequence represents the 45 kDa subunit of the human mature processed heparanase dimer, which is used in

CC the exemplification of the present invention.

XX Sequence 386 AA;

SQ

Query Match 100.0%; Score 82; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 1e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
|||||

Db 62 SWELGNEPNSFLKKA 76

RESULT 10

ADY27061

ID ADY27061 standard; protein; 460 AA.

XX AC

XX ADY27061;

XX

DT 05-MAY-2005 (first entry)

XX

DE Heparanase inhibitor protein #4.

XX

XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;

XX KW neurological disease; viral infection; infection; cytostatic;

XX KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;

XX KW heparanase modulator; enzyme purification.

XX

OS Homo sapiens.

XX

XX WO2005016227-A2.

XX

XX 24-FEB-2005.

XX

XX 12-AUG-2004; 2004WO-IL000744.

XX

XX 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX

XX Van-Gelder JM, Miron D;

XX

XX WPI; 2005-182203/19.

XX

XX Regulating heparanase activity, useful for treating heparanase-associated

XX PT diseases (e.g. cancer, inflammation, cardiovascular diseases,

XX PT neurological diseases or viral diseases) comprises modulating heparanase

XX PT activation.

XX

XX Disclosure; SEQ ID NO 37; 211pp; English.

XX

XX The invention relates to a method of regulating heparanase activity in a

XX CC tissue or regulating a biological process depending at least in part on

XX CC heparanase activity comprising modulating heparanase activation. The

XX CC invention also relates to methods of treating a heparanase- or heparin

XX CC binding protein-associated disease or disorder in a subject, a

XX CC pharmaceutical composition for use in the treatment of a heparanase-

XX CC associated disease or disorder comprising a therapeutic amount of an

XX CC agent capable of modulating heparanase activation and a pharmaceutical

XX CC carrier or diluent, a method of identifying a protease activator of

XX CC heparanase, a protease substrate mimetic comprising a peptide

XX CC representing a subset or all substrate residues or cleavage sites of

XX CC human heparanase or an equivalent non-human heparanase, a method of

XX CC producing active heparanase and a method of modulating an adhesion

XX CC activity of heparanase. The composition and methods are useful for

XX CC modulating heparanase activation and for treating heparanase-associated

XX CC diseases or disorders such as cancer, inflammation, cardiovascular

XX CC diseases, neurological diseases or viral infections. This sequence

XX CC represents a heparanase inhibitor protein used in the scope of the

XX CC invention.

XX

SQ Sequence 460 AA;

Query Match 100.0%; Score 82; DB 9; Length 460;

Best Local Similarity 100.0%; Pred. No. 1.2e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
|||||

Db 136 SWELGNEPNSFLKKA 150

RESULT 11

AEB87589

ID AEB87589 standard; protein; 486 AA.

XX AC

XX AEB87589;

XX

DT 06-OCT-2005 (first entry)

XX

DE Human heparanase 65delta20 deletion mutant.

XX

XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;

XX KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;

XX KW leukemia; lymphoma; dermatological disease; hematological disease;

XX KW immune disorder; inflammation; antiinflammatory; renal disease;

XX KW nephrotropic; endocrine disease; genitourinary disease;

XX KW autoimmune disease; immunosuppressive; drug screening; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

XX WO2005071070-A2.

XX

XX 04-AUG-2005.

XX

XX 20-JAN-2005; 2005WO-IL000068.

XX

XX 22-JAN-2004; 2004IL-00160025.

XX PR 28-JUL-2004; 2004US-00901943.

XX

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX

XX Vlodavsky I, Ilan N, Levy-Adam F;

XX

XX WPI; 2005-564219/57.

XX DR N-PSDB; AEB87588.

XX

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,

XX PT for treating or inhibiting malignant proliferative, inflammatory, kidney

XX PT disorder or autoimmune disorder.

XX

XX Claim 107; SEQ ID NO 31; 167pp; English.

XX

XX The present sequence is that of a deletion mutant of human heparanase,

XX CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the

XX CC native protein. The recombinant protein is deficient of heparanase

XX CC endoglycosidase catalytic activity. The invention relates to amino acid

XX CC sequences derived from the N-terminus region of the 50 kDa subunit of

XX CC heparanase, particularly in the regions between amino acid residues 158-

XX CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise

XX CC heparin-binding domains. The invention also provides an antibody directed

XX CC to these sequences, in particular the 158-171 peptide, and compositions

XX CC and uses of this antibody as a heparanase inhibitor. A screening method

XX CC is provided for specific heparanase inhibitors. Claimed pharmaceutical

XX CC compositions comprising (i) a peptide derived from the N-terminus region

XX CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a

XX CC peptide, or (iii) an antibody which specifically recognizes the peptide

XX CC are used for the inhibition of heparanase catalytic activity associated

XX CC with an inflammatory disorder, kidney disease, autoimmune disease,

XX CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or

XX CC with a malignant proliferative disorder, especially a solid or non-solid

XX CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.

XX

SQ Sequence 486 AA;

Query Match 100.0%; Score 82; DB 9; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
| | | | | | | | | | | | | | | |
Db 184 SWELGNEPNSFLKKA 198

RESULT 12
ADZ18996
ID ADZ18996 standard; protein; 492 AA.
AC
XX
AC ADZ18996;
XX
DT 16-JUN-2005 (first entry)
XX
DE Hep106 construct protein.
XX
XX Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;
KW inflammation; neoplasia; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX WO2005030962-A1.
PN
XX
PD 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
PF
XX 26-SEP-2003; 2003US-0506479P.
PR
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
PA
XX Lahm A, Nardella C, Pallaoaro M, Steinkuhler C;
PI
XX WPI; 2005-273382/28.
DR
XX N-PSDB; ADZ18997.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 17; 65pp; English.
PS
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hep106 construct protein
CC used in the scope of the invention.
XX
XX
SQ Sequence 492 AA;

Query Match 100.0%; Score 82; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
| | | | | | | | | | | | | | | |
Db 168 SWELGNEPNSFLKKA 182

RESULT 13
AEB87562
ID AEB87562 standard; protein; 493 AA.
XX

AC AEB87562;
XX
DT 06-OCT-2005 (first entry)
XX
XX Human heparanase 65delta15 deletion mutant.
XX
KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasia;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; mutein.
XX
XX Homo sapiens.
OS
OS Synthetic.
XX
PN WO2005071070-A2.
XX
XX 04-AUG-2005.
PD
XX 20-JAN-2005; 2005WO-IL000068.
XX
XX 22-JAN-2004; 2004IL-00160025.
PR
XX 28-JUL-2004; 2004US-00901943.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
PA
XX Vlodavsky I, Ilan N, Levy-Adam F;
PI
XX WPI; 2005-564219/57.
DR
XX N-PSDB; AEB87561.
XX
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX
XX Claim 105; SEQ ID NO 4; 167pp; English.
PS
XX
XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
XX
SQ Sequence 493 AA;

Query Match 100.0%; Score 82; DB 9; Length 493;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
| | | | | | | | | | | | | | | |
Db 169 SWELGNEPNSFLKKA 183

RESULT 14
ADZ18999
ID ADZ18999 standard; protein; 495 AA.
XX

AC ADZ18999;
XX
XX
DT 16-JUN-2005 (first entry)
XX
XX
DE Hep109 construct protein.
XX
XX Enzyme engineering; heparanase; hepl109; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
XX Synthetic.
OS
XX WO2005030962-A1.
FN
XX
XX 07-APR-2005.
PD
XX
XX 17-SEP-2004; 2004WO-EP010517.
PF
XX
XX 26-SEP-2003; 2003US-0506479P.
PR
XX 20-JAN-2004; 2004US-0537729P.
PR
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
PA
XX
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
DR
XX N-PSDB; ADZ18998.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 20; 65pp; English.
PS
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepl109 construct protein
CC used in the scope of the invention.
XX
XX
SQ Sequence 495 AA;
Query Match 100.0%; Score 82; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.3e-05; Mismatches 0; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SWELGNPNPSFLKKA 15
Db 171 SWELGNPNPSFLKKA 185
|||||
RESULT 15
AEB87587
ID AEB87587 standard; protein; 497 AA.
XX
AC AEB87587;
XX
XX 06-OCT-2005 (first entry)
DT
XX
XX Human heparanase 65delta10 deletion mutant.
DE
XX
XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; mutain.
XX
XX Homo sapiens.
OS

OS Synthetic.
XX
PN WO2005071070-A2.
XX
XX 04-AUG-2005.
PD
XX
XX 20-JAN-2005; 2005WO-IL000068.
PF
XX
XX 22-JAN-2004; 2004IL-00160025.
PR
XX 28-JUL-2004; 2004US-00901943.
PR
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
PA
XX
XX Vlodavsky I, Ilan N, Levy-Adam F;
XX
XX WPI; 2005-564219/57.
DR
XX N-PSDB; AEB87586.
XX
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
PT
XX
XX Claim 106; SEQ ID NO 29; 167pp; English.
PS
XX
XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
XX
SQ Sequence 497 AA;
Query Match 100.0%; Score 82; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 1.3e-05; Mismatches 0; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SWELGNPNPSFLKKA 15
Db 184 SWELGNPNPSFLKKA 198
|||||
RESULT 16
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX
XX
AC ADZ19000;
XX
XX 16-JUN-2005 (first entry)
DT
XX
XX HepGS3 construct protein.
DE
XX
XX Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
XX Synthetic.
OS
XX WO2005030962-A1.
PN
XX

PD 07-APR-2005.
 XX 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX
 PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 XX
 DR WPI; 2005-273382/28.
 DR N-PSDB; ADZ19001.
 XX
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Example 2; SEQ ID NO 21; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepG3 construct protein
 CC used in the scope of the invention.
 XX
 SQ Sequence 501 AA;
 Query Match 100.0%; Score 82; DB 9; Length 501;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNPNFLKKA 15
 DB 177 SWELGNPNFLKKA 191
 |||||
 RESULT 17
 ADZ19005
 ID ADZ19005 standard; protein; 507 AA.
 XX
 AC ADZ19005;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE HepG6 construct protein.
 XX
 KW Enzyme engineering; heparanase; hepG6; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 XX
 PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 XX
 DR WPI; 2005-273382/28.
 DR N-PSDB; ADZ19003.
 XX

PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Example 2; SEQ ID NO 26; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepG6 construct protein
 CC used in the scope of the invention.
 XX
 SQ Sequence 507 AA;
 Query Match 100.0%; Score 82; DB 9; Length 507;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNPNFLKKA 15
 DB 183 SWELGNPNFLKKA 197
 |||||
 RESULT 18
 ADY27058
 ID ADY27058 standard; protein; 508 AA.
 XX
 AC ADY27058;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 DE Human inactive heparanase protein.
 XX
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.
 XX
 OS Homo sapiens.
 XX
 PN WO2005016227-A2.
 XX
 PD 24-FEB-2005.
 XX
 PF 12-AUG-2004; 2004WO-IL000744.
 XX
 PR 14-AUG-2003; 2003US-0494800P.
 PR 12-JAN-2004; 2004US-0535492P.
 XX
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
 XX
 PI Van-Gelder JM, Miron D;
 XX
 DR WPI; 2005-182203/19.
 XX
 PT Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.
 XX
 PS Claim 257; SEQ ID NO 34; 211pp; English.
 XX
 CC The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an

CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular
 CC diseases, neurological diseases or viral infections. This sequence
 CC represents a human inactive heparanase protein used in the scope of the
 CC invention.
 XX SQ Sequence 508 AA;

Query Match 100.0%; Score 82; DB 9; Length 508;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 184 SWELGNEPNSFLKKA 198

RESULT 19
 ADZ19006
 ID ADZ19006 standard; protein; 526 AA.
 XX
 AC ADZ19006;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE HepHyaluro construct protein.
 XX
 KW Enzyme engineering; heparanase; hepHyaluro; metastasis;
 KW autoimmune disease; inflammation; neoplasm; immune disorder;
 KW antiinflammatory; cytostatic; immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.
 XX
 PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
 XX
 DR WPI; 2005-273382/28.
 DR N-PSDB; ADZ19007.
 XX
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Example 2; SEQ ID NO 27; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepHyaluro construct
 CC protein used in the scope of the invention.
 XX SQ Sequence 526 AA;

Query Match 100.0%; Score 82; DB 9; Length 526;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 202 SWELGNEPNSFLKKA 216

RESULT 20
 ABB07815
 ID ABB07815 standard; protein; 527 AA.
 XX
 AC ABB07815;
 XX
 DT 03-JUL-2002 (first entry)
 XX
 DE Chicken signal peptide/human heparanase chimeric protein sequence.
 XX
 KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.
 XX
 OS Synthetic.
 OS Gallus gallus.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /note="chicken heparanase signal peptide"
 FT Protein 20..527
 FT /note="human heparanase mature protein"
 XX
 PN US2002034810-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 16-AUG-2001; 2001US-00930218.
 XX
 PR 20-SEP-2000; 2000US-00666390.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;
 XX
 DR WPI; 2002-338926/37.
 DR N-PSDB; ABL40753.
 XX
 PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.
 XX
 PS Disclosure; Page 26-28; 39pp; English.
 XX
 CC The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a chicken signal peptide/human heparanase chimeric protein
 CC sequence
 XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 5; Length 527;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 203 SWELGNEPNSFLKKA 217

RESULT 21
 ABW02018
 ID ABW02018 standard; protein; 527 AA.

XX AC ABW02018;
 XX AC

DT 12-FEB-2004 (first entry)

XX Chimeric human-chicken heparanase protein.

KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.

OS Chimeric - Gallus gallus.
 OS Chimeric - Homo sapiens.

PN US2003180788-A1.

XX 25-SEP-2003.

XX 08-MAY-2003; 2003US-00431438.

XX 20-SEP-2000; 2000US-00666390.

PR 16-AUG-2001; 2001US-00930218.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2003-843931/78.

DR N-PSDB; AAD63532.

XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.

XX Example; Page 26-28; Opp; English.

XX The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
 CC some bacterial and protozoa infections. Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamine. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein

XX Sequence 527 AA;

Query Match 100.0%; Score 82; DB 7; Length 527;

Best Local Similarity 100.0%; Pred. No. 1.4e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 203 SWELGNEPNSFLKKA 217

RESULT 22
 ADO63826

XX ADO63826 standard; protein; 527 AA.

AC ADO63826;

XX 26-AUG-2004 (first entry)

XX Chimeric heparanase mutant E343A, SEQ ID:11.

XX Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein.

OS Homo sapiens.

OS Gallus gallus.

OS Synthetic.

OS Chimeric.

XX Key Location/Qualifiers

FT Peptide 1..18

FT /note= "Chicken heparanase signal peptide"

FT Region 19..527

FT /note= "Corresponds to residues 35-543 of human

FT heparanase mutant E343A (SEQ ID NO:8) "

FT Active-site 209

FT /note= "Active site proton donor"

FT Misc-difference 327

FT /note= "Ala replaces wild-type Glu (active site

FT nucleophile). Corresponds to residue 343 of human

FT heparanase mutant E343A (SEQ ID NO:8) "

XX WO2004048558-A2.

XX 10-JUN-2004.

XX 24-NOV-2003; 2003WO-IL000989.

XX 24-NOV-2002; 2002IL-00153059.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;

XX WPI; 2004-450373/42.

DR N-PSDB; ADO63820.

XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.

XX Claim 10; SEQ ID NO 11; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing

CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be use in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase mutant E343A.

XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
 |||||
 Db 203 SWELGNEPNSFLKKA 217

RESULT 23
 ADZ19004
 ID ADZ19004 standard; protein; 527 AA.

XX AC ADZ19004;

XX DT 16-JUN-2005 (first entry)

XX DE HepGS4 construct protein.

XX KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX DR N-PSDB; ADZ19002.

XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.

XX PS Example 2; SEQ ID NO 25; 65pp; English.

XX CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepGS4 construct protein
 CC used in the scope of the invention.

SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 9; Length 527;
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
 |||||
 Db 203 SWELGNEPNSFLKKA 217

RESULT 24

AAV34173

ID AAY34173 standard; protein; 530 AA.

XX AC AAY34173;

XX DT 15-NOV-1999 (first entry)

XX DE Human pre-proheparanase protein sequence.

XX KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;
 KW therapy.

XX OS Homo sapiens.

XX PN WO9943830-A2.

XX PD 02-SEP-1999.

XX PF 18-FEB-1999; 99WO-US001489.

XX PR 24-FEB-1998; 98US-0075706P.

XX PR 26-MAR-1998; 98US-0079401P.

XX PA (PHAA) PHARMACIA & UPJOHN CO.

XX PI Heinrikson RL, Fairbanks MB, Mildner AM;

XX WPI; 1999-540598/45.

XX DR N-PSDB; AAZ11236.

XX PT New isolated platelet heparanase polypeptides, used to develop products
 for, e.g. wound healing and blocking angiogenesis.

XX PS Claim 12; Fig 7; 57pp; English.

XX CC This sequence is the human pre-proheparanase of the invention. This
 CC sequence was isolated from human platelets. The heparanase can be used
 CC for identifying agents which alter heparanase activity. The heparanase
 CC can be used for wound healing or for blocking angiogenesis or
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic
 CC retinopathy or solid tumours, or for the degradation of heparin and the
 CC neutralisation of heparin's anticoagulant properties during surgery.

XX CC Inhibitors of heparanase activity can be used in the treatment of
 CC atherosclerosis, asthma, and other inflammatory diseases, vascular restenosis,
 CC disorders, and central nervous system (CNS) and neurodegenerative
 CC diseases. The products can also be used for detection and diagnosis. The
 CC purified heparanase, both recombinantly produced human heparanase and
 CC heparanase isolated from human platelet activity, allows for the
 CC convenient selection of compounds having anti-heparanase activity, i.e.
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase
 CC activity. Inhibition of heparanase activity can be measured by blocking
 CC heparanase-mediated release of radioactive fragments from in vivo
 CC radiolabelled (HSPG)/heparin

XX SQ Sequence 530 AA;

Query Match 100.0%; Score 82; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
|||
Db 206 SWELGNEPNSFLKKA 220

RESULT 25

AAAY17083
ID AAY17083 standard; protein; 532 AA.

XX AC AAY17083;
XX DT 21-JUL-1999 (first entry)
XX DE
XX OS Homo sapiens.
XX PN WO9921975-A1.
XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
XX KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
XX KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
XX KW human; HSPG.

XX OS Homo sapiens.

XX PN WO9921975-A1.

XX PD 06-MAY-1999.

XX PF 28-OCT-1998; 98WO-AU000898.

XX PR 28-OCT-1997; 97AU-00000062.

XX PR 09-DEC-1997; 97AU-00000812.

XX PA (AUSU) UNIV AUSTRALIAN NAT.

XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

XX DR WPI; 1999-312956/26.

XX DR N-PSDB; AAX37260.

XX PT Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing.

XX PS Claim 6; Page 76-79; 112pp; English.

XX CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG

XX SQ Sequence 532 AA;

Query Match 100.0%; Score 82; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
|||
Db 219 SWELGNEPNSFLKKA 233

RESULT 26

AAAY02345
ID AAY02345 standard; protein; 543 AA.

XX AC AAY02345;
XX DT 09-JUL-1999 (first entry)

XX DE A human heparanase protein.

XX KW Heparanase; hp; modulator; heparin-binding growth factor;
XX KW cellular response; cytokine; cell interaction; plasma lipoprotein;
XX KW cellular susceptibility; infection; disintegration;
XX KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
XX KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
XX KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX OS Homo sapiens.

XX PN WO9911798-A1.

XX PD 11-MAR-1999.

XX PF 31-AUG-1998; 98WO-US017954.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 02-JUL-1998; 98US-00109386.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PA (FRIE/) FRIEDMAN M M.

XX PI Pecker I, Vlodaysky I, Feinstein E;

XX XX WPI; 1999-302255/25.

XX DR N-PSDB; AAX35648.

XX PT New human polynucleotide useful for treating angiogenesis, restenosis,
PT and inflammation.

XX PS Claim 6; Fig 1; 63pp; English.

XX CC The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of heparin-
CC binding growth factors, cellular responses to heparin-binding growth
CC factors and cytokines, cell interaction with plasma lipoproteins
CC cellular susceptibility to viral, protozoal and bacterial infections or
CC disintegration of neurodegenerative plaques. Heparanase may be useful for
CC conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and
CC renal failure in biopsy specimens, plasma samples, and body fluids. The
CC present sequence represents human heparanase

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
|||
Db 219 SWELGNEPNSFLKKA 233

RESULT 27

AAAY17082
ID AAY17082 standard; protein; 543 AA.

XX AC AAY17082;
XX DT 21-JUL-1999 (first entry)

```
XX DE Human heparanase enzyme.
XX OS
XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
XX KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
XX KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
XX KW human; HSPG.
XX OS
XX OS Homo sapiens.
XX PN WO9921975-A1.
XX PD 06-MAY-1999.
XX PF 28-OCT-1998; 98WO-AU000898.
XX PR 28-OCT-1997; 97AU-00000062.
XX PR 09-DEC-1997; 97AU-00000812.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
XX DR N-PSDB; AAX37259.
XX PT Polynucleotides encoding mammalian endoglucuronidases, especially
XX PT heparanases, useful to promote wound healing.
XX PS Claim 6; Page 69-73; 112pp; English.
XX CC The invention relates to nucleic acid sequences that encode heparanase
XX CC enzymes having endoglucuronidase activity. Recombinant heparanases are
XX CC capable of removing the HS side chain from heparan sulfate proteoglycan
XX CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
XX CC inhibit heparanase, this is useful for treatment of a physiological or
XX CC medical condition associated with elevated heparanase activity, such as
XX CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
XX CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
XX CC rat heparanases can be used to enhance wound healing, especially
XX CC associated with tissue development and repair. The conditions mentioned
XX CC above can be diagnosed using specific antibodies, and also using primers
XX CC and probes specific for the heparanase polynucleotides. Other uses of the
XX CC heparanases include sequencing sulfated molecules such as HSPG. The
XX CC present sequence represents a human heparanase
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233

RESULT 28
AAY57590
ID AAY57590 standard; protein; 543 AA.
XX AC AAY57590;
XX DT 02-MAR-2000 (first entry)
XX DE Human heparanase.
XX KW Human; heparanase; hpa; genetic modification; expression; anticancer;
XX KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
XX KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
XX KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
XX KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
XX KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;

KW micrometastasis; autoimmune lesion; kidney failure.
XX OS
XX OS Homo sapiens.
XX PN WO9957244-A1.
XX PD 11-NOV-1999.
XX PF 29-APR-1999; 99WO-US009256.
XX PR 01-MAY-1998; 98US-00071618.
XX PR 02-MAR-1999; 99US-00260038.
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PA (FRIE/) FRIEDMAN M M.
XX PI Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;
XX PI Peleg Y, Shlomi Y;
XX DR WPI; 2000-062144/05.
XX DR N-PSDB; AAZ39195.
XX PT Engineered cells that express recombinant heparanase, useful
XX PT therapeutically, e.g. for treating angiogenesis and to screen for
XX PT specific inhibitors, potential anticancer agents.
XX PS Claim 3; Page 107-109; 118pp; English.
XX CC The present invention describes genetically modified cells (A) containing
XX CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
XX CC and express recombinant heparanase (II). Heparanase cleaves heparan
XX CC sulphate (HS) at specific intrachain sites, resulting in release of
XX CC heparin-binding growth factors, enzymes and proteins that are sequestered
XX CC by HS in basement membranes, extracellular matrix or cell surfaces. It
XX CC may also be implicated in tumour angiogenesis and metastases. (II) is
XX CC potentially useful in wound healing and for treating angiogenesis,
XX CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
XX CC infection and cystic fibrosis. It can also be used to neutralise heparin
XX CC (an alternative to protamine) and to screen for specific inhibitors
XX CC (potentially useful for treating cancer and metastases). Antibodies
XX CC raised against (II) are used for immunodetection and diagnosis of
XX CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
XX CC in large quantities, in a form that is homogeneously processed and
XX CC activated/neutralised by a dedicated protease. The present sequence
XX CC represents human heparanase
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233

RESULT 29
AAB08849
ID AAB08849 standard; protein; 543 AA.
XX AC AAB08849;
XX DT 15-JAN-2001 (first entry)
XX DE Amino acid sequence of a human heparanase polypeptide.
XX KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
XX KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
XX KW wound healing; infection; burn; angiogenesis; restenosis;
XX KW atherosclerosis; inflammation; neurodegenerative disease;
XX KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
```

OS Homo sapiens.
 PN WC200052178-A1.
 XX
 PD 08-SEP-2000.
 PF 14-FEB-2000; 2000WO-US003542.
 XX
 PR 01-MAR-1999; 99US-00258892.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodavsky I, Feinstein E;
 XX
 DR WPI; 2000-579289/54.
 DR N-PSDB; AAA75051.
 XX
 PT New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
 XX
 PS Claim 22; Fig 1; 152pp; English.
 XX
 CC The present sequence represents a human protein with heparanase catalytic
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
 CC particularly in treating tumour, inflammation or autoimmunity.
 CC Particularly, the polynucleotide is useful in modulating the
 CC bioavailability of heparin-binding growth factors, cellular responses to
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
 CC susceptibility to certain viral and some bacterial and protozoa
 CC infections, or disintegration of neurodegenerative plaques. The
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
 CC radiation burns), and in the treatment of angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
 CC bacterial or protozoa infections
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 82; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 219 SWELGNEPNSFLKKA 233
 RESULT 30
 AAY52990
 ID AAY52990 standard; protein; 543 AA.
 XX
 AC AAY52990;
 XX
 XX 21-FEB-2000 (first entry)
 DT
 DE Human heparanase protein sequence.
 XX
 KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
 KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
 KW autoimmune disease; anticancer; kidney disease.
 XX
 OS Homo sapiens.
 XX
 PN WO9957153-A1.
 XX
 PD 11-NOV-1999.

XX 29-APR-1999; 99WO-US009255.
 XX
 PR 01-MAY-1998; 98US-00071739.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
 XX
 XX WPI; 2000-052944/04.
 DR N-PSDB; AAZ33290.
 XX
 PT Heparanase-specific molecular probes useful for diagnosis and treatment,
 PT e.g. of tumors, and for targeted drug delivery.
 XX
 PS Example; Page 81-82; 90pp; English.
 XX
 CC The present invention describes heparanase-specific molecular probes,
 CC useful for methods of detecting heparanase in situ. The probes and anti-
 CC heparanase antibodies are used to detect or quantify the expression of
 CC heparanase, for diagnosis and monitoring of diseases (especially
 CC metastasis), for treatment of heparanase-associated diseases (e.g.
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
 CC sepsis and inflammatory or autoimmune disease), for targeted drug
 CC delivery (e.g. of anticancer agents) and as research reagents. The
 CC present sequence represents human heparanase, which is used in the
 CC exemplification of the present invention
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 82; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 219 SWELGNEPNSFLKKA 233
 RESULT 31
 AAY97635
 ID AAY97635 standard; protein; 543 AA.
 XX
 AC AAY97635;
 XX
 DT 20-APR-2001 (first entry)
 DE
 DE Human heparanase protein sequence.
 XX
 KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200100643-A2.
 XX
 PD 04-JAN-2001.
 PF 19-JUN-2000; 2000WO-IL000358.
 XX
 PR 25-JUN-1999; 99US-0140801P.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Pecker I, Michal I, Itzhaki H;

XX WPI; 2001-137930/14.
XX New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy protocols
PT for angiogenesis, restenosis, atherosclerosis, or inflammation.
XX
PS Disclosure; Page 64-65; 67pp; English.
XX
CC This sequence represents a heparanase of the invention. The heparanase
CC DNA and protein sequences are useful in wound healing, angiogenesis,
CC restenosis, atherosclerosis, inflammation, pulmonary diseases,
CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and
CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
CC sequence is particularly useful in gene therapy
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
|||||
RESULT 32
AAB86206
ID AAB86206 standard; protein; 543 AA.
XX
AC AAB86206;
XX
DT 24-AUG-2001 (first entry)
XX
DE Human heparanase inhibitor protein.
XX
KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral edema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human.
XX
OS Homo sapiens.
XX
FN DE19955803-A1.
XX
PD 23-MAY-2001.
XX
PF 19-NOV-1999; 99DE-01055803.
XX
PR 19-NOV-1999; 99DE-01055803.
XX
PA (KNOL) KNOLL AG.
XX
PI Herr D, Hahn A, Laux V;
XX
DR WPI; 2001-368371/39.
DR N-PSDB; AAH20940.
XX
PT Treatment or prevention of cardiac insufficiency and related conditions,
PT e.g. pulmonary congestion and dyspnoea, comprises administration of
PT heparanase inhibitor.
XX
PS Disclosure; Page 11-13; 16pp; German.
XX
CC This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
CC the invention has cardiant, nephrotropic and hepatotropic activity. The
CC products of the invention can be used in human and veterinary medicine,
CC for the treatment or prevention of congestive heart failure e.g. primary
CC cardiomyopathy. Associated conditions treated or prevented with the
CC inhibitor are especially peripheral odemas, pulmonary and hepatic

CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
CC nocturia can also be treated. This sequence represents the human
CC heparanase protein described in the method of the invention
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
|||||
RESULT 33
AAB88361
ID AAB88361 standard; protein; 543 AA.
XX
AC AAB88361;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0090.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes.
XX
OS Homo sapiens.
XX
FN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-00114090.
XX
PR 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
DR N-PSDB; AAF93788.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development.
XX
PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme

CC linked immunosorbant assay (ELISA). Examples of diseases which may be
XX treated include rheumatoid arthritis and diabetes

SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

|||||

Db 219 SWELGNEPNSFLKKA 233

RESULT 34

ABB07813

ID ABB07813 standard; protein; 543 AA.

XX

AC ABB07813;

XX

DT 03-JUL-2002 (first entry)

XX

DE Human heparanase sequence.

XX

KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..35

FT /note= "signal peptide"

FT Protein 36..543

FT /note= "mature protein"

XX

PN US2002034810-A1.

XX

XX

PD 21-MAR-2002.

XX

PF 16-AUG-2001; 2001US-00930218.

XX

PR 20-SEP-2000; 2000US-00666390.

XX

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA

PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX

DR WPI; 2002-338926/37.

XX

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
PT to treat various heparin-related disorders and the signal peptide is
PT useful in production of membrane-targeted or secreted recombinant
PT proteins.

XX

XX Disclosure; Fig 1a; 39pp; English.

PS

CC The invention relates to an isolated avian and reptile nucleic acid,
CC encoding a polypeptide with heparanase catalytic activity. The signal
CC peptide of the nucleic acid can be used to express membrane-associated or
CC secreted proteins in heterologous expression systems. The encoded
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC invasion, and to intervene with pathologies associated with impaired
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoa and bacterial infections or
CC disintegration of neurodegenerative plaques. The present sequence
CC represents a human heparanase protein sequence used in similarity studies

XX SQ Sequence 543 AA;

Query Match

Best Local Similarity 100.0%; Score 82; DB 5; Length 543;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

|||||

Db 219 SWELGNEPNSFLKKA 233

RESULT 35

ADD18950

ID ADD18950 standard; protein; 543 AA.

XX

AC ADD18950;

XX

DT 15-JAN-2004 (first entry)

XX

DE Human disease related protein SeqID439.

XX

KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
KW antiarteriosclerotic; vulnary; gene therapy;
KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
KW glucose transportation; catecholamine synthesis; iron transport;
KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
KW inflammatory condition; wound healing.

OS Homo sapiens.

XX

XX WO2003018621-A2.

PN

PD 06-MAR-2003.

XX

PF 23-AUG-2002; 2002WO-GB003892.

XX

PR 23-AUG-2001; 2001GB-00020558.

XX

PR 05-OCT-2001; 2001GB-00024037.

XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX

XX WPI; 2003-290046/28.

DR

DR N-PSDB; ADD18951.

XX

PT New substantially purified polypeptide, useful for diagnosing or treating
PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
PT wound healing.

XX

PS Claim 25; SEQ ID NO 439; 424pp; English.

XX

CC This invention relates to novel human genes and gene product which are
CC implicated in certain disease states. Compounds which modulate the
CC proteins of the invention may have cytostatic, antiinflammatory,
CC ophthalmological, antiarteriosclerotic or vulnary activities. The
CC sequences of the invention may be useful for gene therapy. The invention
CC may be useful for diagnosing or treating a hypoxia-regulated condition,
CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
CC erythropoiesis, or the biological response to hypoxia conditions
CC including processes such as glycolysis, gluconeogenesis, glucose
CC transportation, catecholamine synthesis, iron transport or nitric oxide
CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
CC inflammatory conditions or wound healing. The present sequence is that of
CC a disease related protein of the invention.

XX SQ Sequence 543 AA;

Query Match

Best Local Similarity 100.0%; Score 82; DB 7; Length 543;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

|||||

Db 219 SWELGNPNFLKKA 233

RESULT 36
ADG88800
ID ADG88800 standard; protein; 543 AA.
AC ADG88800;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human hpa protein.
XX
KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.
XX
OS Homo sapiens.
XX
PN US2003161823-A1.
XX
PD 28-AUG-2003.
XX
PF 14-JAN-2003; 2003US-00341582.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 06-FEB-2001; 2001US-00776874.
PR 05-SEP-2001; 2001WO-IL000830.
PR 19-NOV-2001; 2001US-00988113.
XX
PA (ILAN/) ILAN N.
PA (VL0D/) VL0DAVSKY I.
PA (YACO/) YACOBY-ZEEVI O.
PA (PECK/) PECKER I.
PA (FEIN/) FEINSTEIN E.
XX
PI Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
XX
DR WPI; 2003-897910/82.
DR N-PSDB; ADG88799, ADG88801, ADG88832.
XX
PT Composition for treating a wound comprising recombinant heparanase is
PT useful to induce or accelerate wound healing and induce or accelerate
PT angiogenesis.
XX
PS Claim 2; SEQ ID NO 10; 143pp; English.
XX
CC The present invention relates to methods and compositions for inducing
CC and/or accelerating wound healing via the catalytic activity of
CC heparanase. The invention is used to induce or accelerate a healing
CC process, particularly of an ulcer, burn, laceration, surgical incision,
CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate
CC angiogenesis. The present sequence is human hpa protein.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SWELGNPNFLKKA 15
Db 219 SWELGNPNFLKKA 233
RESULT 37
ADL16379
ID ADL16379 standard; protein; 543 AA.
AC ADL16379;
XX
DT 06-MAY-2004 (first entry)
XX

DE Human heparanase partial protein.
XX
KW Human; heparanase; heparanase-dependent cancer; cancer;
KW autoimmune reaction; inflammation; chromosome 4; enzyme.
XX
OS Homo sapiens.
XX
PN US2003236215-A1.
XX
PD 25-DEC-2003.
XX
PF 09-JUN-2003; 2003US-00456573.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 08-NOV-1999; 99US-00435739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2004-070610/07.
XX
CC New antisense oligonucleotide hybridizable with a polynucleotide encoding
CC a polypeptide with heparanase activity, useful for treating diseases such
CC as cancer and autoimmune disorders.
XX
PS Claim 3; SEQ ID NO 10; 108pp; English.
XX
CC The invention relates to an antisense oligonucleotide (ASO) comprising a
CC polynucleotide or a polynucleotide analogue of at least 10 bases being
CC hybridisable in vivo, under physiological conditions, with a portion of
CC a polynucleotide strand encoding a polypeptide having heparanase
CC catalytic activity. Also included are a method of in vivo downregulating
CC heparanase activity (comprising administering the ASO in vivo), a method
CC of treating a subject suffering from a pathological condition
CC (characterised by heparanase activity, comprising administering ASO to
CC the subject), a pharmaceutical composition comprising the ASO and a
CC carrier, an antisense nucleic acid construct (comprising a promoter
CC sequence and a polynucleotide sequence directing the synthesis of an
CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,
CC under physiological conditions, with a polynucleotide strand encoding a
CC polypeptide having heparanase catalytic activity), a method of in vivo
CC downregulating heparanase activity (comprising administering in vivo the
CC antisense nucleic acid construct), a pharmaceutical composition
CC comprising the antisense nucleic acid construct and a carrier, and an
CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
CC analogue of at least 10 bases being hybridisable in vivo, under
CC physiological conditions, with a portion of a polynucleotide strand being
CC characterised by forming at least a portion of an untranslated region
CC (UTR) for a polynucleotide strand encoding a polypeptide having
CC heparanase catalytic activity. The methods and compositions of the
CC present invention are useful for the prevention and/or treatment of
CC diseases or conditions associated with aberrant heparanase activity, such
CC as heparanase-dependent cancer, cancer, autoimmune reaction and
CC inflammation. The gene for human heparanase is located on chromosome 4.
XX
CC The present sequence is a human heparanase protein.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SWELGNPNFLKKA 15
Db 219 SWELGNPNFLKKA 233
RESULT 38
ADK52086
ID ADK52086 standard; protein; 543 AA.

```
XX AC ADK52086;
XX DT 20-MAY-2004 (first entry)
XX DE Human atopic dermatitis/psoriasis-associated protein #1.
XX KW Human, atopic dermatitis; psoriasis; dermatological; anti-inflammatory;
XX OS antipsoriatic; rash.
XX PN Homo sapiens.
XX WO2004016785-A1.
XX PD 26-FEB-2004.
XX PF 06-AUG-2003; 2003WO-JP009999.
XX PR 06-AUG-2002; 2002JP-00229319.
XX PR 14-MAY-2003; 2003JP-00136544.
XX PA (GENO-) GENOX RES INC.
XX PA (UYJU-) UNIV JUNTENDO.
XX PI Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
XX PI Mitsuishi K;
XX DR WPI; 2004-214514/20.
XX DR N-PSDB; ADK51968.
XX PT Detecting atopic dermatitis or psoriasis comprises assaying levels of
XX PT expression of an indicator gene at a rash site and non-rash site of a
XX PT person with atopic dermatitis or psoriasis.
XX PS Example 2; SEQ ID NO 119; 484pp; Japanese.
XX CC The invention relates to detecting atopic dermatitis or psoriasis
XX CC comprising assaying the levels of expression of an indicator gene at a
XX CC rash site and non-rash site of a person with atopic dermatitis or
XX CC psoriasis, comparing these levels with those of a healthy person, and
XX CC determining that if the levels of indicators are higher or lower, then
XX CC this indicates the disease. Also included are a reagent for detecting
XX CC atopic dermatitis or psoriasis, a kit for screening for treatments, a
XX CC transgenic non human vertebrate animal models for the diseases, an agent
XX CC for inducing the diseases in mice and a DNA chip for assaying for the
XX CC indicator genes. The method is used for treatment, detection and animal
XX CC models for research of atopic dermatitis and psoriasis. The present
XX CC sequence is a protein encoded by an indicator gene of the invention.
XX SQ Sequence 543 AA;
XX CC Query Match 100.0%; Score 82; DB 8; Length 543;
XX CC Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db |||||
219 SWELGNEPNSFLKKA 233
RESULT 39
ADM48716
ID ADM48716 standard; protein; 543 AA.
XX AC ADM48716;
XX DT 03-JUN-2004 (first entry)
XX DE Human hpa protein #1.
XX KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX KW human.
```

```
XX OS Homo sapiens.
XX PN US2003217375-A1.
XX PD 20-NOV-2003.
XX PF 24-FEB-2003; 2003US-00371218.
XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSCHMIDT O.
XX PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX PI Chajek-Shaul T, Goldshmidt O;
XX DR WPI; 2004-021918/02.
XX DR N-PSDB; ADM48715, ADM48717.
XX PT New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancers, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX PS Example 1; SEQ ID NO 10; 106pp; English.
XX CC The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX SQ Sequence 543 AA;
XX CC Query Match 100.0%; Score 82; DB 8; Length 543;
XX CC Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db |||||
219 SWELGNEPNSFLKKA 233
RESULT 40
ADM48759
ID ADM48759 standard; protein; 543 AA.
XX AC ADM48759;
XX DT 03-JUN-2004 (first entry)
XX DE Human hpa protein #2.
XX KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX KW human.
XX OS Homo sapiens.
```



```
XX PN US2003217375-A1.
XX PD 20-NOV-2003.
XX PF 24-FEB-2003; 2003US-00371218.
XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSCHMIDT O.
XX PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX PI Chajek-Shaul T, Goldshmidt O;
XX DR WPI; 2004-021918/02.
XX DR N-PSDB; ADM48748.
XX
XX New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancer, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
XX Example 10; Fig 16; 106pp; English.
XX
XX The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX
XX SQ Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
XX
RESULT 41
ADN05074
ID ADN05074 standard; protein; 543 AA.
XX
XX ADN05074;
XX
XX 01-JUL-2004 (first entry)
XX
XX Antipsoriatic protein sequence #716.
XX
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
XX
XX 08-APR-2004.
XX
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
XX
RESULT 41
ADN05074
ID ADN05074 standard; protein; 543 AA.
XX
XX ADN05074;
XX
XX 01-JUL-2004 (first entry)
XX
XX Antipsoriatic protein sequence #716.
XX
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
XX
XX 08-APR-2004.
XX
```

```
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX
XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN05073.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX
XX Claim 9; SEQ ID NO 1468; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX
XX SQ Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
XX
RESULT 42
ADN04902
ID ADN04902 standard; protein; 543 AA.
XX
XX ADN04902;
XX
XX 01-JUL-2004 (first entry)
XX
XX Antipsoriatic protein sequence #631.
XX
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
XX
XX 08-APR-2004.
XX
XX 25-SEP-2003; 2003WO-US030907.
XX
XX 25-SEP-2002; 2002US-0414006P.
XX
XX (GETH ) GENENTECH INC.
XX
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX
XX WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX
XX Claim 9; SEQ ID NO 1296; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
```

```

CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
RESULT 43
AD063831
ID AD063831 standard; protein; 543 AA.
XX
AC AD063831;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human heparanase mutant E378A.
XX
KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein; enzyme.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Active-site 225
FT Active-site /note= "Active site proton donor"
FT Active-site 343
FT /note= "Active site nucleophile"
FT Misc-difference 378
FT /note= "Ala replaces wild-type Glu"
XX
PN WO2004048558-A2.
XX
PD 10-JUN-2004.
XX
PF 24-NOV-2003; 2003WO-IL000989.
XX
PR 24-NOV-2002; 2002IL-00153059.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX
WPI; 2004-450373/42.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Example 4; Page; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The
XX invention also relates to the heparanase-derived proteins and host cells
XX comprising the nucleic acid constructs of the invention. The heparanase-
XX derived proteins are especially mutants of human heparanase in which the

```

```

CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (AD063822-AD063824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (AD063825-
CC AD063827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a human
CC heparanase mutant E378A created in an example of the invention which
CC retains its heparanase catalytic activity. The present sequence is not
CC shown in the invention, but is derived from the protein sequence of
CC GenBank accession number AF144325 and the information provided on page
CC 70.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 82; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
Db 219 SWELGNEPNSFLKKA 233
RESULT 44
AD063823
ID AD063823 standard; protein; 543 AA.
XX
AC AD063823;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human heparanase mutant E343A, SEQ ID:8.
XX
KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Active-site 225
FT Active-site /note= "Active site proton donor"
FT Misc-difference 343
FT /note= "Ala replaces wild-type Glu (active site
XX nucleophile)"
XX
PN WO2004048558-A2.
XX
PD 10-JUN-2004.
XX
PF 24-NOV-2003; 2003WO-IL000989.
XX
PR 24-NOV-2002; 2002IL-00153059.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

```

XX WPI; 2004-450373/42.
 DR N-PSDB; ADO63817.
 XX
 PT New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 XX
 XX
 XX Claim 9; SEQ ID NO 8; 128pp; English.
 PS
 XX The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents the human
 CC heparanase mutant E343A.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNPNFLKKA 15
 |||||
 Db 219 SWELGNPNFLKKA 233
 |||||
 RESULT 45
 ADO63832
 ID ADO63832 standard; protein; 543 AA.
 XX
 AC ADO63832;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Human heparanase mutant E396A.
 XX
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnerable; mutant; mutein; enzyme.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Active-site 225
 FT /note= "Active site proton donor"
 FT Active-site 343

FT Misc-difference 396 /note= "Active site nucleophile"
 FT /note= "Ala replaces wild-type Glu"
 XX
 PN WO2004048558-A2.
 PD 10-JUN-2004.
 XX
 XX 24-NOV-2003; 2003WO-IL000989.
 XX
 XX 24-NOV-2002; 2002IL-00153059.
 PR
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
 PI WPI; 2004-450373/42.
 XX
 DR New nucleic acid construct comprising heparanase-derived polypeptide,
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 PT
 PS Example 4; Page; 128pp; English.
 XX
 CC The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a human
 CC heparanase mutant E378A created in an example of the invention which
 CC retains its heparanase catalytic activity. The present sequence is not
 CC shown in the invention, but is derived from the protein sequence of
 CC GenBank accession number AF144325 and the information provided on page
 CC 70.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 SWELGNPNFLKKA 15
 |||||
 Db 219 SWELGNPNFLKKA 233
 |||||
 RESULT 46
 ADO80372
 ID ADO80372 standard; protein; 543 AA.
 XX
 AC ADO80372;
 XX
 DT 21-OCT-2004 (first entry)

```

XX DE Heparanase protein.
XX DE
XX KW cytostatic; epidermal growth factor receptor modulator; identification;
XX KW therapeutic response; cancer; EGFR; biomarker.
XX
XX OS Homo sapiens.
XX
XX PN WO2004063709-A2.
XX
XX PD 29-JUL-2004.
XX
XX PF 08-JAN-2004; 2004WO-US000368.
XX
XX PR 08-JAN-2003; 2003US-0438735P.
XX
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX PI Amler LC, Januario T;
XX
XX DR WPI; 2004-544114/52.
XX DR N-PSDB; ADQ80253.
XX
XX PT Identifying a mammal that will respond therapeutically to a method of
XX PT treating cancer comprises comparing the level of a biomarker in a mammal
XX PT before and after exposure to an epidermal growth factor receptor (EGFR)
XX PT modulator.
XX
XX PS Disclosure; SEQ ID NO 144; 520pp; English.
XX
XX CC The invention relates to a method of identifying a mammal that will
XX CC respond therapeutically to a method of treating cancer by administering
XX CC an epidermal growth factor receptor (EGFR) modulator by comparing the
XX CC level of a biomarker in a mammal before and after exposure to an EGFR
XX CC modulator. The method comprises: (a) measuring, in the mammal, the level
XX CC of at least one biomarker identified in the specification; (b) exposing
XX CC the mammal to the EGFR modulator; and (c) measuring in the mammal the
XX CC level of the biomarker, where a difference in the level in step (c)
XX CC compared to step (a) indicates that the mammal will respond
XX CC therapeutically to the method of treating cancer. The method and
XX CC biomarkers are useful for identifying a mammal that will respond
XX CC therapeutically to a method of treating cancer by administering an
XX CC epidermal growth factor receptor (EGFR) modulator. This sequence
XX CC corresponds to one of the biomarkers whose levels of expression is
XX CC measured in the method of the invention.
XX
XX SQ Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SWELGNEPNFLKKA 15
XX |||||
XX Db 219 SWELGNEPNFLKKA 233
XX
XX RESULT 47
XX ADR88210
XX ID ADR88210 standard; protein; 543 AA.
XX AC
XX AC ADR88210;
XX
XX DT 18-NOV-2004 (first entry)
XX
XX DE Human preproheparanase.
XX
XX CC Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
XX KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
XX KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
XX KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
XX KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
XX KW human; heparanase; enzyme.

```

```

XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..35
XX FT Protein /label= Signal_peptide
XX FT Region /label= Mature_heparanase
XX FT Region 36..109
XX FT Domain /note= "8 KDa subunit of mature heparanase dimer"
XX FT Domain 89..107
XX FT Region /note = Functional peptide epitope
XX FT Domain 158..1543
XX FT Domain /note= "45 KDa subunit of mature heparanase dimer"
XX FT Domain 219..233
XX FT Active-site /note = Functional peptide epitope
XX FT Binding-site /note= "Active site residue"
XX FT Domain 258..266
XX FT Domain /note= "Putative heparin binding domain"
XX FT Domain 294..307
XX FT Domain /note = Functional peptide epitope
XX FT Active-site 334..348
XX FT Binding-site /note = Functional peptide epitope
XX FT Binding-site /note= "Active site residue"
XX FT Domain 414..420
XX FT Domain /note= "Putative heparin binding domain"
XX FT Domain 437..446
XX FT /note = Functional peptide epitope
XX
XX US2004170631-A1.
XX
XX PD 02-SEP-2004.
XX
XX PF 28-NOV-2003; 2003US-00722502.
XX
XX PR 02-SEP-1997; 97US-00922170.
XX PR 01-MAY-1998; 98US-00071739.
XX PR 04-NOV-1998; 98US-00186200.
XX PR 19-FEB-2003; 2003US-00368044.
XX PR 22-AUG-2003; 2003US-00645659.
XX
XX PA (YACO/) YACOBY-ZEEVI O.
XX PA (PERE/) PERETZ T.
XX PA (MIRO/) MIRON D.
XX PA (SHLO/) SHLOMI Y.
XX PA (PECK/) PECKER I.
XX PA (AYAL/) AYAL-HERSHKOVITZ M.
XX PA (FEIN/) FEINSTEIN E.
XX PA (VGEL/) VAN GELDER J M.
XX PA (VLOD/) VLODAVSKY I.
XX PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
XX PI Ayal-HersHKovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
XX PI Friedmann Y;
XX
XX WPI; 2004-625084/60.
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
XX useful for treating heparanase-associated conditions such as inflammation
XX or cancer, comprises administering a drug and an anti-heparanase antibody
XX complex.
XX
XX Claim 2; SEQ ID NO 4; 58pp; English.
XX
XX The invention relates to a method of targeted drug delivery to a tissue
XX of a patient, the tissue expressing heparanase. The method comprises
XX providing a complex of a drug directly or indirectly linked to an anti-
XX heparanase antibody, and administering the complex to the patient. In the
XX targeted drug delivery, the antibody comprises an antibody or its portion
XX capable of specifically binding to at least one epitope of a heparanase

```

CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia (AML), chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelium, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is human preproheparanase.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 219 SWELGNEPNSFLKKA 233

RESULT 48
 ADP25079
 ID ADP25079 standard; protein; 543 AA.

XX ADP25079;

XX 18-NOV-2004 (first entry)

XX PRO polypeptide SEQ ID NO:2257.

XX PRO; antiinflammatory; antiarthritic; antirheumatic; immunosuppressive;
 KW osteopathic; antidiabetic; dermatological; antipsoriatic; antiallergic;
 KW antiasthmatic; hepatotropic; respiratory; gene therapy; immune system.

XX Unidentified.

XX WO2004041170-A2.

XX 21-MAY-2004.

XX 30-OCT-2003; 2003WO-US034312.

XX 01-NOV-2002; 2002US-0423394P.

XX (GETH) GENENTECH INC.

XX Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WI;
 PI Wu TD;

XX WPI: 2004-419628/39.

XX N-PSDB; ADP25078.

XX New PRO polypeptides and polynucleotides, useful for treating e.g.
 PT erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated
 PT renal disease, or demyelinating diseases of the central or peripheral

PT nervous system.

XX Claim 7; SEQ ID NO 2257; 2940pp; English.

XX The invention relates to a novel isolated nucleic acid and the PRO
 CC polypeptide encoded by it. A protein of the invention has
 CC antiinflammatory, antiarthritic, dermatological, antipsoriatic, antiallergic,
 CC osteopathic, antidiabetic, hepatotropic, and respiratory activity. A polynucleotide
 CC of the invention may have a use in gene therapy. The PRO polypeptide, its
 CC agonist, antagonist, or antibody that specifically binds to the
 CC polypeptide is useful for treating an immune related disorder such as
 CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,
 CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an
 CC idiopathic inflammatory myopathy, Sjogren's syndrome, autoimmune
 CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune
 CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal
 CC disease, a demyelinating disease of the central or peripheral nervous
 CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,
 CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary
 CC disease, infectious or autoimmune chronic active hepatitis, primary
 CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,
 CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's
 CC disease, an autoimmune or immune-mediated skin disease, a bullous skin
 CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic
 CC disease, asthma, allergic rhinitis, atopic dermatitis, food
 CC hypersensitivity, urticaria, an immunologic disease of the lung,
 CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity
 CC pneumonitis, a transplantation associated disease, graft rejection or
 CC graft-versus-host disease. The present sequence represents a PRO protein
 CC of the invention.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
 |||||
 Db 219 SWELGNEPNSFLKKA 233

RESULT 49

ADT78177

ID ADT78177 standard; protein; 543 AA.

XX ADT78177;

XX 13-JAN-2005 (first entry)

XX Human heparanase protein.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Binding-site 157..162

FT Binding-site /note="Putative heparin binding site"

FT Binding-site 271..277

FT Binding-site /note="Putative heparin binding site"

FT Binding-site 426..433

FT Binding-site /note="Putative heparin binding site"

XX US2004213789-A1.

XX 28-OCT-2004.

PD

```
XX 22-AUG-2003; 2003US-00645659.
XX
XX 02-SEP-1997; 97US-00922170.
XX 01-MAY-1998; 98US-00071739.
XX 04-NOV-1998; 98US-00186200.
XX 18-FEB-2003; 2003US-00368044.
XX
XX (YACOBY-ZEEVI O.
XX (PERE/) PERETZ T.
XX (MIRO/) MIRON D.
XX (SHLO/) SHLOMI Y.
XX (PECK/) PECKER I.
XX (AYAL/) AYAL-HERSHKOVITZ M.
XX (FEIN/) FEINSTEIN E.
XX (GELD/) GELDER J M V.
XX (VLOD/) VLODAVSKY I.
XX (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
XX Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
XX Friedmann Y;
XX
XX WPI; 2004-774790/76.
XX
XX New neutralizing monoclonal anti-heparanase antibodies, useful for
XX detecting, treating or preventing cancer, inflammatory or autoimmune
XX disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
XX Claim 5; SEQ ID NO 4; 68pp; English.
XX
XX The invention relates to an isolated antibody or antibody portion capable
XX of specifically binding to or elicited by at least one epitope of a
XX heparanase protein, where the heparanase protein is at least 60%
XX homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
XX where at least one epitope comprises a sequence at least 70% homologous
XX to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
XX a hybridoma cell line comprising a cell line for producing the monoclonal
XX antibody, (b) a method for detecting, treating or preventing a
XX pathological condition or a heparanase-related disorder or condition in a
XX subject, (c) a method for monitoring the state of a heparanase-related
XX disorder or condition in a subject, and (d) a pharmaceutical composition
XX comprising the isolated anti-heparanase antibody or antibody portion and
XX a pharmaceutical carrier. The antibody, methods, and composition are
XX useful for detecting, treating, preventing or monitoring a pathological
XX condition, e.g. angiogenesis, cell proliferation, a cancerous condition
XX (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
XX or prostate cancer), minor cell proliferation, invasion of circulating
XX tumour cells, or a metastatic disease, or a heparanase-related disorder
XX or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
XX (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
XX renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
XX nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
XX carcinoma) in a mammal. This sequence represents human heparanase.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SWELGNBPNSFLKKA 15
XX | | | | | | | | | | | | | | |
XX Db 219 SWELGNBPNSFLKKA 233
XX
XX RESULT 50
XX ADY27036
XX ID ADY27036 standard; protein; 543 AA.
XX AC
XX ADY27036;
XX
XX DT 05-MAY-2005 (first entry)
```

```
XX Human heparanase protein.
XX
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
XX neurological disease; viral infection; infection; cytostatic;
XX antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
XX protease; enzyme; enzyme purification.
XX
XX Homo sapiens.
XX OS
XX WO2005016227-A2.
XX PN
XX 24-FEB-2005.
XX PD
XX 12-AUG-2004; 2004WO-IL000744.
XX PF
XX 14-AUG-2003; 2003US-0494800P.
XX PR
XX 12-JAN-2004; 2004US-0535492P.
XX PR
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PA
XX Van-Gelder JM, Miron D;
XX PI
XX WPI; 2005-182203/19.
XX DR
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
XX diseases (e.g. cancer, inflammation, cardiovascular diseases,
XX PT neurological diseases or viral diseases) comprises modulating heparanase
XX PT activation.
XX
XX Disclosure; SEQ ID NO 8; 211pp; English.
XX
XX The invention relates to a method of regulating heparanase activity in a
XX tissue or regulating a biological process depending at least in part on
XX heparanase activity comprising modulating heparanase activation. The
XX invention also relates to methods of treating a heparanase- or heparin
XX binding protein-associated disease or disorder in a subject, a
XX pharmaceutical composition for use in the treatment of a heparanase-
XX associated disease or disorder comprising a therapeutic amount of an
XX agent capable of modulating heparanase activation and a pharmaceutical
XX carrier or diluent, a method of identifying a protease activator of
XX heparanase, a protease substrate mimetic comprising a peptide
XX representing a subset or all substrate residues or cleavage sites of
XX human heparanase or an equivalent non-human heparanase, a method of
XX producing active heparanase and a method of modulating an adhesion
XX activity of heparanase. The composition and methods are useful for
XX modulating heparanase activation and for treating heparanase-associated
XX diseases or disorders such as cancer, inflammation, cardiovascular
XX diseases, neurological diseases or viral infections. This sequence
XX represents a human heparanase protein used in the scope of the invention.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 9; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SWELGNBPNSFLKKA 15
XX | | | | | | | | | | | | | | |
XX Db 219 SWELGNBPNSFLKKA 233
XX
XX Search completed: June 5, 2006, 12:42:03
XX Job time : 136.384 secs
```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 16.6438 Seconds
(without alignments)
86.714 Million cell updates/sec

Title: US-10-645-659A-8

Perfect score: 82

Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	54.9	648	2 A32576	beta-glucuronidase
2	43	52.4	624	2 T49366	myocyte-specific e
3	42	51.2	114	1 CCR7P	cytochrome c2 - Rh
4	42	51.2	848	2 S48273	probable transcript
5	41.5	50.6	352	2 T08469	endo-1,4-beta-xyla
6	41	50.0	194	2 B83007	conserved hypothet
7	41	50.0	274	2 T19330	hypothetical prote
8	41	50.0	550	2 B84900	hypothetical prote
9	41	50.0	648	2 A25047	beta-glucuronidase
10	40.5	49.4	79	1 C36955	flagellar biosynth
11	40	48.8	117	2 B97391	hypothetical prote
12	40	48.8	117	2 AC2609	hypothetical prote
13	40	48.8	270	2 T32412	hypothetical prote
14	40	48.8	331	2 S54263	rep A protein - Ba
15	40	48.8	413	2 B27539	variant surface gl
16	40	48.8	505	2 A84709	cinnamate-4-hydrox
17	40	48.8	640	2 F81191	hypothetical prote
18	40	48.8	651	2 A26581	beta-glucuronidase
19	40	48.8	716	1 A35269	translation initia
20	40	48.8	1641	2 D82704	conserved hypothet
21	39	47.6	94	2 S48138	ubiquinol-cytochro
22	39	47.6	263	2 T22536	hypothetical prote
23	39	47.6	325	2 A84326	hypothetical prote
24	39	47.6	366	2 T04763	chitinase homolog
25	39	47.6	442	2 F70348	mannanase, probabl
26	39	47.6	459	2 F70189	rod shape-determin
27	39	47.6	473	1 T40391	phosphoprotein pho
28	39	47.6	481	2 H84619	hypothetical prote
29	39	47.6	545	2 S67621	hypothetical prote

30	39	47.6	554	2 H95922	hypothetical membr
31	39	47.6	628	2 AD2408	hypothetical prote
32	39	47.6	651	2 AD3057	glycogen debranchi
33	39	47.6	651	2 B98229	glycogen debranchi
34	39	47.6	660	2 I51684	epithelial sodium
35	39	47.6	669	2 D72278	endo-1,4-beta-mann
36	39	47.6	941	2 S29043	cellulase (EC 3.2.
37	38.5	47.0	642	2 E72215	oligopeptide ABC t
38	38	46.3	100	2 G96948	uncharacterized sm
39	38	46.3	129	2 H87375	hypothetical prote
40	38	46.3	135	2 F86239	protein F20824.4 [
41	38	46.3	139	2 AH1586	bacteriophage phi-
42	38	46.3	217	2 F82788	thymidylate kinase
43	38	46.3	228	2 AE2956	haloacid dehalogen
44	38	46.3	228	2 A98327	hypothetical prote
45	38	46.3	230	2 S49780	hypothetical prote
46	38	46.3	247	2 F70710	probable 3-oxoacyl
47	38	46.3	277	2 T38320	hypothetical prote
48	38	46.3	322	2 T24356	hypothetical prote
49	38	46.3	344	2 S09283	fructose-bisphosph
50	38	46.3	379	2 T04762	chitinase homolog
51	38	46.3	398	2 T04761	chitinase homolog
52	38	46.3	411	2 D86153	hypothetical prote
53	38	46.3	415	2 AG2044	hypothetical prote
54	38	46.3	464	2 T28662	hypothetical prote
55	38	46.3	581	2 B86318	protein F15H18.9 [
56	38	46.3	600	2 AE2570	hypothetical prote
57	38	46.3	661	2 S75005	sensory transducti
58	38	46.3	726	2 S37964	beta-adaptin homol
59	38	46.3	843	2 F96537	hypothetical prote
60	38	46.3	1140	2 S73786	hypothetical prote
61	38	46.3	1540	2 T45619	hypothetical prote
62	38	46.3	2828	2 S59413	hypothetical prote
63	38	46.3	3351	2 T13812	probable membrane
64	37.5	45.7	501	2 A21159	lipophorin - fruit
65	37	45.1	122	2 H82231	cell surface anti
66	37	45.1	128	2 S75259	hypothetical prote
67	37	45.1	143	2 B95408	hypothetical prote
68	37	45.1	146	2 F90141	probable DNA-bind
69	37	45.1	232	2 A27503	hypothetical prote
70	37	45.1	250	2 B40626	testis-specific pr
71	37	45.1	266	2 T05471	probable 2,3-dihyd
72	37	45.1	275	2 T40345	hypothetical prote
73	37	45.1	278	2 AF3430	probable autophago
74	37	45.1	283	2 T50431	ribonuclease III [
75	37	45.1	308	2 T03862	CBP3-like protein
76	37	45.1	310	2 S60425	hypothetical prote
77	37	45.1	342	2 AD2306	probable membrane
78	37	45.1	345	2 F86944	hypothetical prote
79	37	45.1	353	2 AB2483	hypothetical prote
80	37	45.1	363	2 A81016	transcription regu
81	37	45.1	395	2 G84259	histidinol-phospha
82	37	45.1	430	2 S61118	DIN7 protein - Yea
83	37	45.1	437	2 H70918	probable PE protei
84	37	45.1	442	2 F83367	probable glycosyl
85	37	45.1	446	2 T27696	hypothetical prote
86	37	45.1	466	2 S39494	glutathione-disulf
87	37	45.1	472	2 T51429	hypothetical prote
88	37	45.1	477	2 H86466	protein F23M19.7 [
89	37	45.1	499	2 T03213	probable ribosomal
90	37	45.1	521	2 T45608	hypothetical prote
91	37	45.1	538	2 T03232	probable ribosomal
92	37	45.1	541	2 G90516	ABC transporter at
93	37	45.1	564	2 F82194	hypothetical prote
94	37	45.1	631	2 S70910	transferrin-bindin
95	37	45.1	654	2 F83260	hypothetical prote
96	37	45.1	655	2 H82938	probable ABC subst
97	37	45.1	780	2 H84453	probable heat choc
98	37	45.1	781	2 S49340	heat-shock protein
99	37	45.1	804	2 B89961	leucyl-tRNA synthe
100	37	45.1	819	2 S15169	ferric-pseudobacti
101	37	45.1	844	2 C28667	DNA mismatch repai
102	37	45.1	857	2 E98107	DNA mismatch repai

103	37	45.1	1075	2	T47603	beta Galactosidase	176	35	42.7	251	2	C71879	hypothetical prote
104	37	45.1	1150	2	T15277	hypothetical prote	177	35	42.7	254	2	AE0940	probable ribulose-
105	37	45.1	1661	2	H71439	hypothetical prote	178	35	42.7	255	2	H70982	probable nei prote
106	36.5	44.5	458	2	F98271	mraA protein [impo	179	35	42.7	261	2	F90506	hypothetical prote
107	36.5	44.5	458	2	AI3012	phosphoglucumutase	180	35	42.7	272	2	AH2168	hypothetical prote
108	36.5	44.5	475	2	AB3235	phosphomannomutase	181	35	42.7	282	2	I50996	connexin 32.7 - At
109	36	43.9	35	2	E95098	hypothetical prote	182	35	42.7	294	2	T00104	probable dTDP-4-de
110	36	43.9	138	2	S74872	hypothetical prote	183	35	42.7	298	2	B35272	osteoinductive fac
111	36	43.9	138	2	S77514	hypothetical prote	184	35	42.7	298	2	JC4130	osteoinductive fac
112	36	43.9	138	2	S75128	hypothetical prote	185	35	42.7	299	2	A35272	osteoinductive fac
113	36	43.9	154	1	JC1292	protein-export pro	186	35	42.7	314	2	T28879	hypothetical prote
114	36	43.9	173	2	JN0515	fliv protein - Sal	187	35	42.7	328	2	T24006	hypothetical prote
115	36	43.9	203	2	H90287	hypothetical prote	188	35	42.7	331	2	T35809	hypothetical prote
116	36	43.9	209	2	D90788	probable chaperone	189	35	42.7	341	2	T22057	hypothetical prote
117	36	43.9	229	2	E85648	probable chaperone	190	35	42.7	341	2	T26920	hypothetical prote
118	36	43.9	235	2	H71615	probable membrane	191	35	42.7	348	2	E87466	transcription regu
119	36	43.9	259	2	T19091	hypothetical prote	192	35	42.7	365	2	B3829	glycine oxidase (s
120	36	43.9	272	2	T48567	ABA-responsive pro	193	35	42.7	367	2	T36278	hypothetical prote
121	36	43.9	281	2	F84313	hypothetical prote	194	35	42.7	375	2	T11020	transaldolase - My
122	36	43.9	288	2	G64146	yabH protein homol	195	35	42.7	379	2	G70918	hypothetical prote
123	36	43.9	315	2	AF1393	glycosyl transfera	196	35	42.7	392	2	B69522	conserved hypotet
124	36	43.9	315	2	AI1768	glycosyl transfera	197	35	42.7	392	2	H69147	LPS biosynthesis R
125	36	43.9	326	2	S54267	repA protein - Bac	198	35	42.7	395	2	A86458	probable elicitor
126	36	43.9	335	2	T19662	hypothetical prote	199	35	42.7	398	2	B86353	protein F282.6 [im
127	36	43.9	362	2	D82644	sugar-phosphate de	200	35	42.7	399	2	H86453	avr9 homolog F9L11
128	36	43.9	369	2	T04323	mannan endo-1,4-be	201	35	42.7	403	2	A84592	(1-4)-beta-mannan
129	36	43.9	373	2	C70917	probable tal prote	202	35	42.7	406	2	T04817	hypothetical prote
130	36	43.9	414	2	S59783	hypothetical prote	203	35	42.7	425	2	E71075	probable proton gl
131	36	43.9	416	2	A97093	probable membrane	204	35	42.7	431	2	T09048	probable mannan en
132	36	43.9	418	2	S55018	CEN-1 protein - Ca	205	35	42.7	432	2	S51901	tryptophan-trNA 11
133	36	43.9	422	2	S49012	yolk protein 3 - b	206	35	42.7	432	2	AC2146	periplasmic sugar-
134	36	43.9	448	2	T48214	endo-1,4-beta-mann	207	35	42.7	440	2	JC8032	PU.1-binding prote
135	36	43.9	484	2	AF3107	xylokinnase [impo	208	35	42.7	444	2	C81054	phosphoglucumutase
136	36	43.9	488	2	F98179	D-arabinitol kinas	209	35	42.7	445	2	T39415	probable transcrip
137	36	43.9	519	2	T46241	hypothetical prote	210	35	42.7	446	2	H86899	Glutamate-ammonia
138	36	43.9	525	2	AF3601	periplasmic oligop	211	35	42.7	448	2	D97927	glutamate-ammonia
139	36	43.9	528	2	D87443	alkaline phosphata	212	35	42.7	448	2	C95058	glutamine syntheta
140	36	43.9	552	1	KJHUGU	N-acetylglucosamin	213	35	42.7	455	2	S16559	cellulase (EC 3.2.
141	36	43.9	554	2	T50118	kinesin-related pr	214	35	42.7	458	2	S75114	hypothetical prote
142	36	43.9	629	2	S31174	hypothetical prote	215	35	42.7	465	2	H69120	hypothetical prote
143	36	43.9	662	2	B72114	DNA ligase CP0624	216	35	42.7	465	2	C86274	hypothetical prote
144	36	43.9	662	2	E86509	DNA ligase [impor	217	35	42.7	484	2	T24238	hypothetical prote
145	36	43.9	741	2	A27892	translation initia	218	35	42.7	504	2	G82631	glutamine syntheta
146	36	43.9	746	2	C95110	competence protein	219	35	42.7	509	1	S62899	cytochrome P450 (C
147	36	43.9	746	2	A97979	competence protein	220	35	42.7	511	1	VGBE1K	glycoprotein C - h
148	36	43.9	807	2	E90523	leucyl-trna synthe	221	35	42.7	511	1	VGBEF4	hypothetical prote
149	36	43.9	899	2	G36812	hypothetical prote	222	35	42.7	512	2	T13611	hypothetical prote
150	36	43.9	931	2	G69615	ATP-dependent DNA	223	35	42.7	512	2	AC3203	IS3 family transpo
151	36	43.9	989	2	T47503	hypothetical prote	224	35	42.7	512	2	AD2835	IS3 family transpo
152	36	43.9	1084	2	H86229	hypothetical prote	225	35	42.7	512	2	AD3049	IS3 family transpo
153	36	43.9	1124	1	FKPUZ	phytochrome - zucc	226	35	42.7	512	2	G97612	probable transposa
154	36	43.9	1169	2	G72571	probable DNA-direc	227	35	42.7	512	2	G98236	probable transposa
155	36	43.9	1184	2	H86190	hypothetical prote	228	35	42.7	512	2	AC0871	probable integrase
156	36	43.9	1274	2	T04018	hypothetical prote	229	35	42.7	551	2	H86260	protein T1C24.30
157	36	43.9	1832	2	AC2594	glutamate synthase	230	35	42.7	567	2	E91095	type III secretion
158	36	43.9	1858	2	D97376	hypothetical prote	231	35	42.7	567	2	AE5941	type III secretion
159	36	43.9	1939	2	AF0095	probable sideropho	232	35	42.7	584	2	S40013	hypothetical prote
160	35.5	43.3	217	2	B83334	reverse transcript	233	35	42.7	584	2	A82913	glucan 1,4 alpha g
161	35.5	43.3	470	2	E97331	HFI1 protein - yea	234	35	42.7	622	2	G90250	probable formyl tr
162	35.5	43.3	488	2	S61010	hypothetical prote	235	35	42.7	667	2	AD0295	trophozoite antige
163	35.5	43.3	726	2	T17630	hypothetical prote	236	35	42.7	668	2	A44863	UDP-glucose 4-epime
164	35	42.7	108	2	AB1605	B. subtilis comG o	237	35	42.7	689	1	S29621	probable chitinase
165	35	42.7	154	2	A56652	hypothetical prote	238	35	42.7	699	2	AC0504	hypothetical prote
166	35	42.7	176	2	A56652	interleukin-7 prec	239	35	42.7	706	2	T15920	probable membrane
167	35	42.7	176	2	JC4035	interleukin 7 prot	240	35	42.7	732	2	T14233	NADH2 dehydrogenas
168	35	42.7	177	2	C95922	hypothetical prote	241	35	42.7	732	2	T13814	NADH2 dehydrogenas
169	35	42.7	208	2	T12985	hypothetical prote	242	35	42.7	737	2	T14232	NADH2 dehydrogenas
170	35	42.7	217	2	S62795	probable lipoprote	243	35	42.7	750	2	A97501	topoisomerase iv c
171	35	42.7	219	2	H97110	phosphate uptake r	244	35	42.7	750	2	AE2719	topoisomerase iv s
172	35	42.7	230	2	D71601	probable secreted r	245	35	42.7	805	2	B35047	zinc finger protei
173	35	42.7	247	2	AC2786	SEC-independent pr	246	35	42.7	830	2	B44439	protein kinase (EC
174	35	42.7	247	2	E97565	hypothetical prote	247	35	42.7	839	2	A47007	adenylyl cyclase t
175	35	42.7	251	2	F64636	conserved hypotet	248	35	42.7				

249	35	42.7	856	2	A44439	protein kinase (EC	322	34	41.5	248	2	T04758	hypothetical prote
250	35	42.7	857	2	A42861	protein kinase TTK	323	34	41.5	252	2	D64564	hypothetical prote
251	35	42.7	861	1	S34730	1,4-alpha-glucan b	324	34	41.5	262	2	B71947	hypothetical prote
252	35	42.7	1002	2	B91180	probable oxidoredu	325	34	41.5	261	2	H71144	hypothetical prote
253	35	42.7	1002	2	C86026	probable oxidoredu	326	34	41.5	269	2	T44866	hypothetical prote
254	35	42.7	1022	2	T24663	hypothetical prote	327	34	41.5	283	2	H70104	chemotaxis protein
255	35	42.7	1044	2	T43155	nitrite reductase	328	34	41.5	288	2	A75638	probable transposa
256	35	42.7	1044	2	T43160	nitrite reductase	329	34	41.5	288	2	C75426	probable transposa
257	35	42.7	1087	2	T16876	hypothetical prote	330	34	41.5	293	2	T32892	hypothetical prote
258	35	42.7	1115	2	T19137	hypothetical prote	331	34	41.5	297	2	S65185	ribosomal protein
259	35	42.7	1121	2	T138127	phosphoprotein pho	332	34	41.5	304	2	T15729	hypothetical prote
260	35	42.7	1122	1	FKMUA	phytochrome A - Ar	333	34	41.5	305	2	B33282	DNA-binding protei
261	35	42.7	1122	2	D86229	phytochrome A [imp	334	34	41.5	311	2	D69229	conserved hypotet
262	35	42.7	1134	2	A41350	adenylate cyclase	335	34	41.5	319	2	T42423	PAS7 protein homol
263	35	42.7	1166	1	E85151	hypothetical 126K	336	34	41.5	322	2	AD3488	transporter BME118
264	35	42.7	1232	2	T43027	neural cell adhesi	337	34	41.5	327	2	E75618	probable transposa
265	35	42.7	1244	2	A05218	hypothetical prote	338	34	41.5	327	2	C75556	probable transposa
266	35	42.7	1259	2	S25954	gene atpA intron 2	339	34	41.5	327	2	C75624	probable transposa
267	35	42.7	1274	2	T16251	hypothetical prote	340	34	41.5	327	2	B75620	probable transposa
268	35	42.7	1305	2	H41662	150K mating aggreg	341	34	41.5	327	2	A75631	probable transposa
269	35	42.7	1324	2	S51622	cut3 protein - fis	342	34	41.5	327	2	A75633	probable transposa
270	35	42.7	1480	2	T21911	hypothetical prote	343	34	41.5	336	2	T26993	hypothetical prote
271	35	42.7	1483	2	T21914	hypothetical prote	344	34	41.5	340	2	T32891	hypothetical prote
272	35	42.7	1483	2	T21912	hypothetical prote	345	34	41.5	344	2	D70576	probable fructose
273	35	42.7	1510	2	C84727	probable glucan sy	346	34	41.5	344	2	H70030	conserved hypotet
274	35	42.7	1729	2	T18396	erythrocyte membra	347	34	41.5	350	2	A95383	probable ABC trans
275	35	42.7	1878	2	E86189	hypothetical prote	348	34	41.5	354	2	S15660	(2'-5')oligo(A) sy
276	35	42.7	3191	2	T22945	hypothetical prote	349	34	41.5	358	2	S44233	strp protein - str
277	35	42.7	3300	2	D70575	probable ppg prote	350	34	41.5	366	2	T34662	probable integral
278	34.5	42.1	59	2	T26375	hypothetical prote	351	34	41.5	368	2	T47894	actin-like protein
279	34.5	42.1	115	2	T07576	hypothetical prote	352	34	41.5	375	2	JX0131	cellulase (EC 3.2.
280	34.5	42.1	150	2	T10641	carbonic anhydrase	353	34	41.5	383	2	F72352	hypothetical prote
281	34.5	42.1	201	1	A60598	smooth muscle prot	354	34	41.5	385	2	G69101	mannose-1-phosphat
282	34.5	42.1	201	2	A57015	Gene SM22 alpha pr	355	34	41.5	387	2	T22511	hypothetical prote
283	34.5	42.1	372	2	G96973	probable membrane	356	34	41.5	387	2	A86456	unknown protein [i
284	34.5	42.1	398	2	C81729	Mtr/TnaB/TyO perm	357	34	41.5	398	2	JQ0366	phospholipase C (E
285	34.5	42.1	415	2	C71467	probable tyrosine	358	34	41.5	398	2	B49231	phospholipase C, a
286	34.5	42.1	596	2	S32802	apolipoprotein B -	359	34	41.5	398	2	A30565	phospholipase C (E
287	34.5	42.1	904	2	S51299	hypothetical prote	360	34	41.5	422	2	T01197	pectin acetyl ester
288	34.5	42.1	1448	2	T08526	DNA primase traC2	361	34	41.5	426	1	T49035	acid phosphatase (
289	34.5	42.1	1448	2	S37669	traC-2 protein - E	362	34	41.5	443	2	AB1405	probable phosphoes
290	34.5	42.1	1687	2	T39072	DNA2-NAM7 helicase	363	34	41.5	443	2	AI1780	probable phosphoes
291	34.5	42.1	2054	2	T07584	hypothetical prote	364	34	41.5	448	2	TS1400	hypothetical prote
292	34	41.5	40	2	A29184	vitellogenin - tur	365	34	41.5	474	2	T26990	hypothetical prote
293	34	41.5	116	2	C59100	hypothetical prote	366	34	41.5	475	2	AI0010	probable membrane
294	34	41.5	122	2	S32574	serum amyloid prot	367	34	41.5	477	2	T01202	probable RNA helic
295	34	41.5	132	2	A44864	fatty acid-binding	368	34	41.5	504	2	TS2171	cytochrome P450 mo
296	34	41.5	134	2	E83549	hypothetical prote	369	34	41.5	505	2	G75197	lacZ expression re
297	34	41.5	138	2	F75056	hypothetical prote	370	34	41.5	505	2	F71230	hypothetical prote
298	34	41.5	138	2	E71175	hypothetical prote	371	34	41.5	507	2	E90539	hypothetical prote
299	34	41.5	154	2	H89975	hypothetical prote	372	34	41.5	519	2	S45723	P60 protein - oat
300	34	41.5	156	1	A55446	protein-tyrosine-p	373	34	41.5	530	2	JC7979	cellobiohydrolase
301	34	41.5	159	2	H88003	protein K11B4.2 [i	374	34	41.5	563	2	A72300	beta-glucuronidase
302	34	41.5	161	2	F83297	hypothetical prote	375	34	41.5	573	2	T49315	receptor lectin ki
303	34	41.5	162	2	S73299	tryptophan-rich pr	376	34	41.5	574	2	S50756	beta-D-glucosidase
304	34	41.5	170	2	H96931	phycocyanin alpha	377	34	41.5	615	2	S38088	hypothetical prote
305	34	41.5	194	2	S75290	hypothetical prote	378	34	41.5	629	2	T47798	receptor lectin ki
306	34	41.5	194	2	S76919	hypothetical prote	379	34	41.5	636	2	S47299	gene F protein - r
307	34	41.5	200	1	A26694	smooth muscle prot	380	34	41.5	651	2	FC1123	hypothetical prote
308	34	41.5	201	1	JN0774	smooth muscle prot	381	34	41.5	653	2	T34356	hypothetical prote
309	34	41.5	201	1	JS0774	smooth muscle prot	382	34	41.5	656	2	S76505	hypothetical prote
310	34	41.5	210	2	T03865	hypothetical prote	383	34	41.5	666	2	T43171	cytoplasmic signal
311	34	41.5	211	2	A80935	probable TetR-fami	384	34	41.5	667	2	E86728	NADH dehydrogenase
312	34	41.5	215	2	AG0476	TetR-family transc	385	34	41.5	672	2	T24507	hypothetical prote
313	34	41.5	223	2	AH0890	disulfide isomeras	386	34	41.5	692	2	B70484	conserved hypotet
314	34	41.5	224	2	AH1815	two-component resp	387	34	41.5	748	2	TS1738	RNA helicase RH3 [
315	34	41.5	233	1	F69178	conserved hypotet	388	34	41.5	749	2	G86186	hypothetical prote
316	34	41.5	234	2	F65203	hypothetical 26.6	389	34	41.5	783	2	JC5467	cellulase (EC 3.2.
317	34	41.5	234	2	F91240	hypothetical prote	390	34	41.5	785	2	G96825	hypothetical prote
318	34	41.5	234	2	C86088	hypothetical prote	391	34	41.5	785	2	T52059	ent-kaurane syntha
319	34	41.5	239	2	AC2745	glycerophosphoryl	392	34	41.5	800	2	A29003	cellulase (EC 3.2.
320	34	41.5	246	2	T37473	transcription regu	393	34	41.5	800	2	T23780	hypothetical prote
321	34	41.5	246	2	B97526	hypothetical prote	394	34	41.5	822	2	T41941	glycoprotein B - h

395	34	41.5	822	2	JT0611	cellulase (EC 3.2.	468	33	40.2	268	2	AG2603	components of type
396	34	41.5	824	2	JC7532	cellulase (EC 3.2.	469	33	40.2	268	2	AG7385	cpaB protein (Af22
397	34	41.5	825	2	JS0174	cellulase (EC 3.2.	470	33	40.2	270	2	AF2170	inositol monophosph
398	34	41.5	839	2	T21207	hypothetical prote	471	33	40.2	272	2	G71115	hypothetical prote
399	34	41.5	866	2	I79267	trab protein - Bsc	472	33	40.2	275	2	G87125	probable TetR-fami
400	34	41.5	888	2	E82885	hypothetical prote	473	33	40.2	276	2	AD1765	hypothetical cell
401	34	41.5	898	2	H84701	probable villin li	474	33	40.2	277	2	AD1390	hypothetical cell
402	34	41.5	899	2	T42976	hypothetical prote	475	33	40.2	283	2	E90571	conserved hypothet
403	34	41.5	1097	2	T45622	hypothetical prote	476	33	40.2	287	2	S71279	L-ascorbate peroxi
404	34	41.5	1184	1	A34795	kinesin-related pr	477	33	40.2	288	2	S60438	protein N-acetyltr
405	34	41.5	1216	2	F88473	protein F40H6.5 li	478	33	40.2	291	2	AH1680	methylntransferase
406	34	41.5	1229	2	H84465	hypothetical prote	479	33	40.2	292	2	E88482	protein COSD11.9 l
407	34	41.5	1239	2	T42020	class IV chitin sy	480	33	40.2	296	2	I40818	phosphotransbutyry
408	34	41.5	1250	2	T40062	probable nuclear e	481	33	40.2	301	2	E97278	phosphate butyrylt
409	34	41.5	1253	2	T14349	Shyc protein - mou	482	33	40.2	310	2	G02309	UDP-glucuronosyltr
410	34	41.5	1323	2	T11661	phosphoribosylform	483	33	40.2	311	2	I64082	panthothenate kinas
411	34	41.5	1339	2	JC5508	DNA-directed DNA p	484	33	40.2	311	2	F98245	n-formylglutamate
412	34	41.5	1451	2	S42167	190K protein - hum	485	33	40.2	312	2	T02406	hypothetical prote
413	34	41.5	1666	2	A48594	skelemin - mouse	486	33	40.2	314	2	A99446	ABC transporter, A
414	34	41.5	1876	2	T13801	phosphoinositide 3	487	33	40.2	317	2	JC5696	prolyl aminopeptid
415	34	41.5	2358	2	T39569	probable alpha-glu	488	33	40.2	319	2	T38533	sur1 protein homol
416	34	41.5	2371	2	T43432	alpha-glucan synth	489	33	40.2	330	1	JN0561	urukinase-type pla
417	34	41.5	3573	2	S23070	erythronolide synt	490	33	40.2	330	2	D95362	probable AraC-fami
418	34	41.5	5032	1	A35041	ryanodine receptor	491	33	40.2	337	1	E70191	conserved hypothet
419	34	41.5	5037	2	B35041	microtubule-associ	492	33	40.2	340	2	T26923	hypothetical prote
420	34	41.5	5327	2	T13564	hypothetical prote	493	33	40.2	348	1	S32521	alcohol dehydrogen
421	33.5	40.9	85	2	JQ0278	hypothetical prote	494	33	40.2	348	2	C83571	probable binding p
422	33.5	40.9	85	2	S58627	hypothetical prote	495	33	40.2	351	2	T26918	hypothetical prote
423	33.5	40.9	88	2	T43580	type III secretion	496	33	40.2	353	2	AF2560	hypothetical prote
424	33.5	40.9	146	2	S54789	superoxide dismuta	497	33	40.2	355	2	AF2867	endo-1,4-beta-xyla
425	33.5	40.9	159	2	A90546	hypothetical prote	498	33	40.2	357	2	F82878	XAA-PRO aminopept
426	33.5	40.9	174	2	B75114	probable NADH dehy	499	33	40.2	358	2	H36891	transfer complex p
427	33.5	40.9	175	2	T14999	hypothetical prote	500	33	40.2	359	2	E72378	sugar ABC transpor
428	33.5	40.9	259	2	T20399	hypothetical prote	501	33	40.2	365	2	E97644	endo-1,4-beta-xyla
429	33.5	40.9	289	2	AD2519	hypothetical prote	502	33	40.2	368	2	A85768	partial beta-D-glu
430	33.5	40.9	332	2	G90291	endoglucanase prec	503	33	40.2	370	2	D90919	beta-D-glucuronida
431	33.5	40.9	334	2	G90360	endoglucanase prec	504	33	40.2	376	2	G84404	hypothetical prote
432	33.5	40.9	486	2	A69442	conserved hypothet	505	33	40.2	377	2	T06192	probable endo-1,4-
433	33.5	40.9	648	2	T47895	hypothetical prote	506	33	40.2	379	2	A49679	exo-alpha-sialidas
434	33.5	40.9	837	2	A34898	granulocyte colony	507	33	40.2	381	2	H75173	udp-n-acetylglucos
435	33.5	40.9	891	2	AC0149	DNA topoisomerase	508	33	40.2	381	2	C96657	hypothetical prote
436	33.5	40.9	940	2	T00056	hypothetical prote	509	33	40.2	388	1	WMLJBT	membrane-bound lyt
437	33.5	40.9	1662	2	T18540	moA protein precu	510	33	40.2	390	2	AH0125	hypothetical prote
438	33	40.2	35	2	T07870	major latex protei	511	33	40.2	394	2	A86431	nuclear protein EM
439	33	40.2	51	2	S78688	hypothetical prote	512	33	40.2	397	2	T09579	prescorin-8Y methy
440	33	40.2	109	2	T00234	hypothetical prote	513	33	40.2	398	2	T44688	alkane 1-monooxyge
441	33	40.2	114	1	CCRF2P	cytochrome c2 - Rh	514	33	40.2	401	1	A31266	trp-Asp repeat pro
442	33	40.2	124	2	B47098	binding-protein-de	515	33	40.2	404	2	T40553	DNA polymerase pro
443	33	40.2	147	2	A70462	ribosomal protein	516	33	40.2	406	2	T42561	cellulase (EC 3.2.
444	33	40.2	151	2	H64104	dupP diphosphatase	517	33	40.2	409	2	B25156	cellulase (EC 3.2.
445	33	40.2	169	2	H83972	hypothetical prote	518	33	40.2	427	2	T06191	actin like protein
446	33	40.2	173	2	G81428	hypothetical prote	519	33	40.2	427	2	T06195	probable endo-1,4-
447	33	40.2	177	2	C71329	hypothetical prote	520	33	40.2	429	2	T06195	probable endo-1,4-
448	33	40.2	185	2	T36546	hypothetical prote	521	33	40.2	430	2	T14417	S-locus-specific g
449	33	40.2	189	2	S41410	insecticynanin prec	522	33	40.2	430	2	T14420	S-locus-specific g
450	33	40.2	190	2	H71245	hypothetical prote	523	33	40.2	441	2	AB1290	ATP-dependent RNA
451	33	40.2	196	2	A39021	pinlin - Haemophilu	524	33	40.2	442	2	AH1661	ATP-dependent RNA
452	33	40.2	211	2	T38645	hypothetical prote	525	33	40.2	451	2	F87039	protoporphyrinogen
453	33	40.2	211	2	T38645	conserved hypothet	526	33	40.2	463	2	T16218	translation elonga
454	33	40.2	216	2	A60331	pinlin precursor -	527	33	40.2	466	2	C42360	cellulase (EC 3.2.
455	33	40.2	222	2	C69326	conserved hypothet	528	33	40.2	471	1	A35867	cytochrome P450 71
456	33	40.2	227	2	A48412	hypothetical prote	529	33	40.2	471	2	T47568	fructokinase-like
457	33	40.2	236	2	I67432	BCL-2 - rat (fragm	530	33	40.2	473	2	S76985	hypothetical prote
458	33	40.2	236	2	A60190	hypothetical prote	531	33	40.2	474	2	S56748	glutathione syntha
459	33	40.2	237	2	T05249	hypothetical prote	532	33	40.2	478	2	C83739	hypothetical prote
460	33	40.2	238	2	AD2034	hypothetical prote	533	33	40.2	486	2	I40548	bifunctional cellu
461	33	40.2	246	2	E70861	hypothetical prote	534	33	40.2	488	2	F86763	amino acid permeas
462	33	40.2	253	1	VHV01V	nucleocapsid prote	535	33	40.2	488	2	A25156	cellulase (EC 3.2.
463	33	40.2	253	2	S47480	chlorophyll a/b-b1	536	33	40.2	489	1	A53766	phosphoinositide-s
464	33	40.2	253	2	T04642	hypothetical prote	537	33	40.2	494	2	G04961	hypothetical prote
465	33	40.2	254	2	H97081	proline/glycine be	538	33	40.2	494	2	T64382	acetylactate synth
466	33	40.2	255	2	S52341	LHCl-680, photosys	539	33	40.2	498	2	T50525	cytochrome P450 mo
467	33	40.2	266	2	AF3040	conserved hypothet	540	33	40.2	499	2	JN0111	cellulase (EC 3.2.

541	33	40.2	499	2	A27198	cellulase (EC 3.2.	614	33	40.2	992	2	T39315	hypothetical prote
542	33	40.2	501	2	G86460	probable cytochrom	615	33	40.2	1050	2	S54640	KCS1 protein - yea
543	33	40.2	502	1	S45039	cytochrome P450	616	33	40.2	1053	2	T07965	reverse transcript
544	33	40.2	502	2	T52256	cytochrome P-450LX	617	33	40.2	1057	2	T16676	hypothetical prote
545	33	40.2	502	2	T07141	cytochrome P450 CY	618	33	40.2	1063	2	A53164	glutamate dehydrog
546	33	40.2	504	2	H83791	propionyl-CoA carb	619	33	40.2	1103	2	T13590	distal tail fiber
547	33	40.2	504	2	D86332	hypothetical prote	620	33	40.2	1130	2	A48843	MHC class II trans
548	33	40.2	504	2	S54744	cellulase (EC 3.2.	621	33	40.2	1238	2	T32625	hypothetical prote
549	33	40.2	505	2	S39962	endoglucanase - Br	622	33	40.2	1243	2	S07278	tail fiber protein
550	33	40.2	508	2	A1453	hypothetical prote	623	33	40.2	1247	2	T45743	hypothetical prote
551	33	40.2	508	2	G69593	cellulase (EC 3.2.	624	33	40.2	1252	2	D71810	probable type II D
552	33	40.2	508	2	A26874	cellulase (EC 3.2.	625	33	40.2	1294	2	T04278	hypothetical prote
553	33	40.2	510	2	T07119	cytochrome P450 Cp	626	33	40.2	1342	1	S32680	DNA-directed RNA p
554	33	40.2	512	2	T00605	probable cytochrom	627	33	40.2	1364	1	AJF8PP	phosphoribosylamin
555	33	40.2	512	2	T00870	probable cytochrom	628	33	40.2	1465	2	T30891	PH3 protein - mai
556	33	40.2	516	2	D84563	single-stranded-DN	629	33	40.2	1590	2	B87754	protein C43E11.3 (
557	33	40.2	516	2	B71946	probable single-str	630	33	40.2	1601	2	AE2011	hypothetical prote
558	33	40.2	532	2	T10624	reticuline oxidase	631	33	40.2	1622	2	D86428	glutathione S-conj
559	33	40.2	545	2	A10265	periplasmic oligop	632	33	40.2	1623	2	T01369	ABC transporter At
560	33	40.2	547	2	JQ0356	cellulase (EC 3.2.	633	33	40.2	1631	1	SAZQK1	major merozoite su
561	33	40.2	547	2	T44743	probable thiamin b	634	33	40.2	1639	2	S05603	major merozoite su
562	33	40.2	548	2	S33788	Photinus-luciferin	635	33	40.2	1640	2	A24594	probable major sur
563	33	40.2	560	2	C84632	hypothetical prote	636	33	40.2	1662	2	T01893	hypothetical prote
564	33	40.2	576	2	T48585	auxin-regulated pr	637	33	40.2	1674	2	G96736	hypothetical prote
565	33	40.2	579	2	JT0494	alpha-glucosidase	638	33	40.2	1726	1	SAZQGM	major merozoite su
566	33	40.2	583	2	T25690	hypothetical prote	639	33	40.2	1726	2	A45948	major merozoite su
567	33	40.2	594	2	C70021	butyryl-CoA dehydr	640	33	40.2	1739	2	A48298	sodium channel hom
568	33	40.2	600	2	E53290	oligopeptide trans	641	33	40.2	1786	1	MMMSB1	laminin beta-1 cha
569	33	40.2	603	1	GBEC9C	beta-glucuronidase	642	33	40.2	1993	2	AF1450	probable peptidogl
570	33	40.2	606	2	H97012	hypothetical prote	643	33	40.2	2004	2	D88948	protein ZK1005.1 (
571	33	40.2	620	2	T30765	hypothetical prote	644	33	40.2	2047	2	S53611	MI8P1 protein - ra
572	33	40.2	631	2	C90552	lipoprotein (impor	645	33	40.2	2437	2	D84904	hypothetical prote
573	33	40.2	636	2	G97148	molybdopter bios	646	33	40.2	2697	2	T25444	hypothetical prote
574	33	40.2	640	2	G96034	conserved hypotet	647	33	40.2	3738	2	T05501	hypothetical prote
575	33	40.2	641	2	JC7331	gamma-glutamyltran	648	33	40.2	5005	2	F82884	phage terminase-li
576	33	40.2	649	2	T47609	hypothetical prote	649	32.5	39.6	151	2	H97133	hypothetical prote
577	33	40.2	652	2	A42245	thiamin biosynthes	650	32.5	39.6	201	2	T22600	hypothetical prote
578	33	40.2	659	2	A45184	protein-tyrosine k	651	32.5	39.6	235	2	T04451	hypothetical prote
579	33	40.2	659	2	I49553	protein-tyrosine k	652	32.5	39.6	239	2	S65825	hypothetical prote
580	33	40.2	665	2	C71667	propionyl-CoA carb	653	32.5	39.6	250	2	E75376	2-hydroxyhepta-2,4
581	33	40.2	665	2	G97819	hypothetical prote	654	32.5	39.6	254	2	H83334	probable transcript
582	33	40.2	666	2	T35864	hypothetical prote	655	32.5	39.6	263	2	D86644	hypothetical prote
583	33	40.2	668	2	S55023	brown protein - fr	656	32.5	39.6	274	2	A72241	endoglucanase - Th
584	33	40.2	677	2	A32611	beta-galactosidase	657	32.5	39.6	282	2	G95130	transcription regu
585	33	40.2	684	2	T47694	probable serine/th	658	32.5	39.6	332	2	S27726	hypothetical prote
586	33	40.2	711	2	A37051	outer membrane/ph	659	32.5	39.6	335	2	JT0569	chondromodulin-I p
587	33	40.2	719	2	JC1200	alpha-glucosidase	660	32.5	39.6	381	2	F81073	succinyl-diaminopi
588	33	40.2	720	2	T08838	piIQ protein - Nei	661	32.5	39.6	382	2	S32148	exo-alpha-sialidas
589	33	40.2	732	2	B84902	hypothetical prote	662	32.5	39.6	382	2	S01339	exo-alpha-sialidas
590	33	40.2	733	2	G87060	conserved membrane	663	32.5	39.6	450	2	F65183	4-alpha-l-fucosylt
591	33	40.2	739	2	H72364	aspartokinase II -	664	32.5	39.6	452	2	AB0921	probable 4-alpha-l
592	33	40.2	741	2	A26572	bsg25D protein - f	665	32.5	39.6	478	1	C42790	cystathionine beta
593	33	40.2	742	2	F84643	hypothetical prote	666	32.5	39.6	504	2	D71615	hypothetical prote
594	33	40.2	756	2	AB1088	chitinase B homolo	667	32.5	39.6	551	1	A55760	cystathionine beta
595	33	40.2	756	2	AB1452	chitinase B homolo	668	32.5	39.6	561	1	A42790	cystathionine beta
596	33	40.2	758	2	T51335	subtilisin-like pr	669	32.5	39.6	661	1	G71063	probable ferrous i
597	33	40.2	780	2	B70112	DNA mismatch repai	670	32.5	39.6	661	2	B75084	ferrous iron trans
598	33	40.2	782	2	A82940	hypothetical prote	671	32.5	39.6	770	1	TWBVA2	transcription fact
599	33	40.2	829	2	T19514	hypothetical prote	672	32.5	39.6	1220	2	A56136	jagged protein pre
600	33	40.2	860	2	T37768	probable vacuolar	673	32.5	39.6	1290	2	A56493	leucocyte common a
601	33	40.2	864	2	A81000	DNA mismatch repai	674	32.5	39.6	1898	2	S46216	hypothetical prote
602	33	40.2	864	2	C82019	DNA mismatch repai	675	32	39.0	68	2	G81063	leucocyte antigen
603	33	40.2	876	1	ITKBAP	DNA topoisomerase	676	32	39.0	99	2	A32716	lg heavy chain C r
604	33	40.2	878	2	A41055	ecdysone receptor	677	32	39.0	102	2	S19464	hypothetical prote
605	33	40.2	880	1	S41420	valine-tRNA ligase	678	32	39.0	103	2	A64012	hypothetical prote
606	33	40.2	884	2	H83322	hypothetical prote	679	32	39.0	110	2	H75048	DNA-directed RNA p
607	33	40.2	932	2	F69552	leucyl-tRNA synthe	680	32	39.0	111	2	S50521	hypothetical prote
608	33	40.2	938	2	T01809	hypothetical prote	681	32	39.0	111	2	B32476	hypothetical prote
609	33	40.2	955	2	E84845	probable villin 2	682	32	39.0	122	2	B66570	U14 ribosomal prot
610	33	40.2	968	2	T01733	hypothetical prote	683	32	39.0	122	2	G72054	ribosomal protein
611	33	40.2	972	1	GNXSIV	genome polyprotein	684	32	39.0	126	2	AG8828	holo-lacyl-carrier
612	33	40.2	972	2	T09624	genome polyprotein	685	32	39.0	130	2	AH2424	hypothetical prote
613	33	40.2	976	2	T50669	villin 2 [imported	686	32	39.0	135	2	B72750	hypothetical prote

687	32	39.0	138	2	AH1214	hypotheical prote	760	32	39.0	315	2	AH1480	transcription regu
688	32	39.0	141	2	H70658	hypotheical prote	761	32	39.0	317	2	S51572	moCA protein - Rhl
689	32	39.0	141	2	JQ1669	hypotheical 16.1K	762	32	39.0	317	2	G72416	sugar ABC transpor
690	32	39.0	151	2	D81814	hypotheical prote	763	32	39.0	317	2	D95204	conserved hypothe
691	32	39.0	155	2	E64344	hypotheical prote	764	32	39.0	317	2	E98071	conserved hypothe
692	32	39.0	155	2	E69808	protein-tyrosine p	765	32	39.0	320	2	E71696	rare lipoprotein A
693	32	39.0	165	2	T46052	ADP-ribosylation f	766	32	39.0	324	2	B66819	oxidoreductase YpJ
694	32	39.0	171	2	T44537	hypotheical prote	767	32	39.0	325	2	D95845	conserved hypothe
695	32	39.0	173	2	AJ0132	transformation com	768	32	39.0	328	2	T20220	phage related prot
696	32	39.0	173	2	H70133	hypotheical prote	769	32	39.0	329	2	F97045	hypotheical prote
697	32	39.0	177	2	F71237	hypotheical prote	770	32	39.0	331	2	T01824	hypotheical prote
698	32	39.0	181	2	T08793	hypotheical prote	771	32	39.0	333	2	B47677	hypotheical prote
699	32	39.0	182	2	E81440	probable acetyltra	772	32	39.0	334	2	C71718	hypotheical prote
700	32	39.0	187	2	T05608	hypotheical prote	773	32	39.0	336	2	C97190	hypotheical prote
701	32	39.0	190	2	T46018	ADP-ribosylation f	774	32	39.0	337	2	C71003	hypotheical prote
702	32	39.0	190	2	H85357	hypotheical prote	775	32	39.0	338	2	G71220	hypotheical prote
703	32	39.0	195	2	AE2755	conserved hypothe	776	32	39.0	339	2	H75187	hypotheical prote
704	32	39.0	203	2	T02868	probable GTP-bind	777	32	39.0	340	2	F70852	virulence-regulati
705	32	39.0	205	2	S71584	GTP-binding protei	778	32	39.0	341	2	T40424	hypotheical prote
706	32	39.0	207	2	T17852	hypotheical prote	779	32	39.0	345	2	F97194	N-acetyl-gamma-glu
707	32	39.0	208	2	T04884	hypotheical prote	780	32	39.0	346	2	H72237	hypotheical prote
708	32	39.0	212	2	D71140	hypotheical prote	781	32	39.0	348	2	T29515	hypotheical prote
709	32	39.0	213	2	A41789	glutathione transf	782	32	39.0	348	2	C70415	cation efflux syst
710	32	39.0	214	2	D59108	hypotheical prote	783	32	39.0	350	2	A48421	ornithine transcar
711	32	39.0	215	2	F90312	conserved hypothe	784	32	39.0	352	2	JC5388	replication initia
712	32	39.0	216	2	T27851	hypotheical prote	785	32	39.0	355	2	A48358	ORF355 - Bradyrhiz
713	32	39.0	216	2	T29039	hypotheical prote	786	32	39.0	358	2	A83249	probable initiatio
714	32	39.0	217	2	H64186	hypotheical prote	787	32	39.0	359	2	T42087	probable 6-phospho
715	32	39.0	218	2	D97536	hypotheical prote	788	32	39.0	360	2	G86435	protein F17F8.7 [i
716	32	39.0	220	1	QQVZC9	F9 protein - sheep	789	32	39.0	363	2	T26172	hypotheical prote
717	32	39.0	223	2	T46018	hypotheical prote	790	32	39.0	365	2	A75375	UDPgalactopyranose
718	32	39.0	230	2	T48569	hypotheical prote	791	32	39.0	367	1	I69653	probable cystathio
719	32	39.0	230	2	A86644	hypotheical prote	792	32	39.0	371	2	B71104	hypotheical prote
720	32	39.0	233	2	S72449	poly(beta-D-mannur	793	32	39.0	371	2	T02102	hypotheical prote
721	32	39.0	236	2	AG2905	hypotheical prote	794	32	39.0	377	2	JC7977	membrane-associat
722	32	39.0	236	2	H97680	probable ATP-bind	795	32	39.0	388	2	T21082	hypotheical prote
723	32	39.0	240	2	AC1269	uroporphyrinogen I	796	32	39.0	389	2	E86634	hypotheical prote
724	32	39.0	240	2	AE1631	uroporphyrinogen I	797	32	39.0	397	2	D91090	probable resistanc
725	32	39.0	246	2	D96491	uroporphyrinogen I	798	32	39.0	397	2	G85935	probable resistanc
726	32	39.0	251	2	T04866	hypotheical prote	799	32	39.0	397	2	D65066	hypotheical prote
727	32	39.0	261	2	S78245	thiamin biosynthes	800	32	39.0	398	2	T46312	hypotheical prote
728	32	39.0	265	2	A72756	probable multiple	801	32	39.0	409	2	F90825	probable integrase
729	32	39.0	272	2	T31191	hypotheical prote	802	32	39.0	411	2	H83788	involved in spore
730	32	39.0	273	2	AB2811	conserved hypothe	803	32	39.0	412	2	C96816	hypotheical prote
731	32	39.0	273	2	E97589	hypotheical prote	804	32	39.0	414	2	B96808	protein F28K19.2 [
732	32	39.0	275	2	T37304	ras GTPase-activat	805	32	39.0	415	2	E83377	probable alcohol d
733	32	39.0	276	2	AE1141	hypotheical prote	806	32	39.0	416	2	A85684	probable integrase
734	32	39.0	276	2	C82084	conserved hypothe	807	32	39.0	418	1	A53888	thermolabile hemol
735	32	39.0	278	2	T29994	hypotheical prote	808	32	39.0	418	2	H82485	NADH dehydrogenase
736	32	39.0	279	2	F86842	prephenate dehydra	809	32	39.0	419	2	D86728	hypotheical prote
737	32	39.0	279	2	S52582	prephenate dehydra	810	32	39.0	421	2	T04753	hypotheical prote
738	32	39.0	281	2	T065530	probable imidazole	811	32	39.0	423	2	H83603	hypotheical prote
739	32	39.0	284	1	F64338	agmatinase (EC 3.5	812	32	39.0	424	2	E87558	cytochrome P450 fa
740	32	39.0	290	2	H64431	glycosyl transfera	813	32	39.0	424	2	H89308	protein F14H8.1 [i
741	32	39.0	293	2	B69866	transcription regu	814	32	39.0	424	2	S17571	carboxypeptidase T
742	32	39.0	294	2	D90194	hypotheical prote	815	32	39.0	428	2	C75110	translation elonga
743	32	39.0	294	2	T45662	hypotheical prote	816	32	39.0	428	2	B75051	hypotheical prote
744	32	39.0	295	2	AD1890	hypotheical prote	817	32	39.0	428	2	G71023	translation elonga
745	32	39.0	300	2	F83804	cation efflux syst	818	32	39.0	433	2	T39745	hypotheical prote
746	32	39.0	301	2	JW0078	amine sulfotransfe	819	32	39.0	439	2	H83699	sodium-dependent c
747	32	39.0	302	2	AH0191	probable polysacch	820	32	39.0	443	2	T08136	probable omega-6 d
748	32	39.0	302	2	H83478	probable binding p	821	32	39.0	445	2	S73996	MG148 homolog Vxps
749	32	39.0	302	2	S27846	hypotheical prote	822	32	39.0	445	2	A80359	hypotheical prote
750	32	39.0	304	2	AG1837	WD-40 repeat prote	823	32	39.0	448	2	D85362	hypotheical prote
751	32	39.0	307	2	A71057	probable sulfatase	824	32	39.0	450	2	AF0987	glutathione-dsulf
752	32	39.0	308	2	T21874	hypotheical prote	825	32	39.0	453	2	E96651	protein T3P18.13 [
753	32	39.0	309	1	B43331	sulfur oxygenase/r	826	32	39.0	454	2	JC4886	bleomycin hydrolas
754	32	39.0	309	2	T20518	hypotheical prote	827	32	39.0	456	2	S61971	hypotheical prote
755	32	39.0	310	2	B65094	hypotheical prote	828	32	39.0	458	2	F86433	protein T17H7.5 [i
756	32	39.0	310	2	F85966	probable transcrip	829	32	39.0	459	2	A13384	biotin carboxylase
757	32	39.0	310	2	G91121	probable transcrip	830	32	39.0	459	2	S13064	1D-myo-inositol-tr
758	32	39.0	311	2	C72381	hypotheical prote	831	32	39.0	461	2	JN0129	1D-myo-inositol-tr
759	32	39.0	314	2	I50811	MHC class I protei	832	32	39.0	464	2	T01957	hypotheical prote

833	32	39.0	466	2	S62330	beta-fructofuranos	906	32	39.0	690	2	E86442	probable PPR-repea
834	32	39.0	470	2	T49272	hypothetical prote	907	32	39.0	692	2	A97013	hypothetical prote
835	32	39.0	471	2	T20938	hypothetical prote	908	32	39.0	693	2	T00256	hypothetical prote
836	32	39.0	477	2	A56449	protoporphyrinogen	909	32	39.0	694	2	I40866	exo-alpha-stalidas
837	32	39.0	481	2	D64883	Aminobenzoyl-gluta	910	32	39.0	695	2	H86900	DNA mismatch repai
838	32	39.0	481	2	A90869	hypothetical prote	911	32	39.0	698	2	T21781	hypothetical prote
839	32	39.0	481	2	H85749	hypothetical prote	912	32	39.0	701	2	S64599	probable membrane
840	32	39.0	486	2	T15281	hypothetical prote	913	32	39.0	702	2	S12638	transposition prot
841	32	39.0	488	2	T242038	catalase [EC 1.11.	914	32	39.0	721	2	T45495	probable transposa
842	32	39.0	491	2	JC7169	54K polar flagella	915	32	39.0	727	2	AH2134	DNA topoisomerase
843	32	39.0	491	2	E95356	probable ABC trans	916	32	39.0	731	1	S20687	DNA ligase (NAD) (
844	32	39.0	492	1	S68856	cytochrome P450 2L	917	32	39.0	746	2	A84800	hypothetical prote
845	32	39.0	492	2	A61382	phosphorylation re	918	32	39.0	751	2	T27691	hypothetical prote
846	32	39.0	493	2	E35115	anthranilate synth	919	32	39.0	752	2	S51866	HPRL protein - yea
847	32	39.0	493	2	T29030	hypothetical prote	920	32	39.0	755	2	G86469	protein Fl2k21.6 (
848	32	39.0	495	2	S43294	bone morphogenetic	921	32	39.0	755	2	A81436	probable outer mem
849	32	39.0	498	2	T51430	dolichyl-phosphate	922	32	39.0	761	2	T41304	probable rna-bindi
850	32	39.0	500	2	D87541	beta-xylosidase [i	923	32	39.0	763	2	I50807	complement factor
851	32	39.0	504	2	C85485	probable carnitine	924	32	39.0	767	1	CO2PCD	cdc10 start contro
852	32	39.0	504	2	C90634	probable carnitine	925	32	39.0	772	2	S62481	hypothetical prote
853	32	39.0	504	2	H64724	probable carnitine	926	32	39.0	775	2	T40847	ubiquitin carboxyl
854	32	39.0	505	2	I38396	protein-tyrosine k	927	32	39.0	777	2	T38769	hypothetical prote
855	32	39.0	505	2	A39128	anthranilate synth	928	32	39.0	786	2	D75630	glycerophosphoryl
856	32	39.0	505	2	B84831	hypothetical prote	929	32	39.0	789	2	B87461	polycerophosphate kina
857	32	39.0	512	2	H96759	probable steroid 2	930	32	39.0	790	2	T19040	hypothetical prote
858	32	39.0	512	2	I49552	protein-tyrosine k	931	32	39.0	794	2	T52441	hypothetical prote
859	32	39.0	512	2	H86206	hypothetical prote	932	32	39.0	797	2	D86247	hypothetical prote
860	32	39.0	515	2	S02194	DNA-directed RNA p	933	32	39.0	808	2	T51138	probable glutamate
861	32	39.0	516	2	G70149	hypothetical prote	934	32	39.0	823	2	B70203	ATP-dependent heli
862	32	39.0	517	2	T02403	probable beta-gluc	935	32	39.0	824	2	T20351	hypothetical prote
863	32	39.0	518	2	S42387	MIPP protein homol	936	32	39.0	827	2	F72414	ribonucleotide red
864	32	39.0	522	2	D72349	conserved hypotet	937	32	39.0	829	2	G86763	DNA topoisomerase
865	32	39.0	522	2	S63057	hypothetical prote	938	32	39.0	867	2	E86815	ClpB protein [mpo
866	32	39.0	531	2	F69949	conserved hypotet	939	32	39.0	874	2	B86322	FlpA1.8 protein -
867	32	39.0	542	1	QREBOA	oligopeptide-bindi	940	32	39.0	878	2	A83748	endo-beta-N-acetyl
868	32	39.0	542	2	B82910	CTP synthetase U02	941	32	39.0	899	2	B87553	DNA topoisomerase
869	32	39.0	543	1	F64871	oligopeptide-bindi	942	32	39.0	908	2	JN0819	transferrin-bindin
870	32	39.0	543	2	F85704	hypothetical prote	943	32	39.0	911	2	T08108	nitrate reductase
871	32	39.0	543	2	G30846	hypothetical prote	944	32	39.0	913	2	T18503	hypothetical prote
872	32	39.0	545	1	S44486	indole-3-pyruvate	945	32	39.0	917	2	E96807	nitrate reductase
873	32	39.0	556	2	T38479	myb-like DNA-bindi	946	32	39.0	917	2	S35228	nitrate reductase
874	32	39.0	557	2	JC5487	cellulase [EC 3.2.	947	32	39.0	918	2	T38786	translation initiat
875	32	39.0	562	2	AF0852	secretory protein	948	32	39.0	920	2	T18852	hypothetical prote
876	32	39.0	563	1	CZCLEM	cellulase [EC 3.2.	949	32	39.0	924	2	S34926	hypothetical prote
877	32	39.0	563	2	B70682	probable nitrite r	950	32	39.0	942	2	T39624	6-phosphofructokin
878	32	39.0	573	2	B70047	two-component sens	951	32	39.0	947	2	H85088	hypothetical prote
879	32	39.0	579	2	H88478	protein F47D12.7 [952	32	39.0	949	2	S54020	probable membrane
880	32	39.0	582	2	AG0650	periplasmic oligop	953	32	39.0	955	2	T21612	hypothetical prote
881	32	39.0	583	2	S50959	probable membrane	954	32	39.0	956	2	T19046	ras GTPase-activat
882	32	39.0	583	2	H69165	hypothetical prote	955	32	39.0	963	2	S45167	chitin synthase (E
883	32	39.0	585	1	A41292	glutamate decarbox	956	32	39.0	971	2	A35697	transcription fact
884	32	39.0	585	1	JH0423	glutamate decarbox	957	32	39.0	987	2	H81722	polymorphic membra
885	32	39.0	585	1	S38533	glutamate decarbox	958	32	39.0	996	2	D86872	beta-galactosidase
886	32	39.0	585	2	JC4064	glutamate decarbox	959	32	39.0	1000	2	F70368	beta-galactosidase
887	32	39.0	586	2	AH2133	ATP-binding protei	960	32	39.0	1007	2	T31333	beta-galactosidase
888	32	39.0	587	2	C86744	myosin-crossreacti	961	32	39.0	1007	2	A30093	beta-galactosidase
889	32	39.0	591	2	S51303	hypothetical prote	962	32	39.0	1013	2	G71460	probable outer mem
890	32	39.0	599	2	H72336	conserved hypotet	963	32	39.0	1015	2	I39697	beta-galactosidase
891	32	39.0	604	2	T37870	RNA-binding / Ran	964	32	39.0	1019	2	C96519	probable disease r
892	32	39.0	606	2	D86434	probable PPR-repea	965	32	39.0	1024	1	GBEC	beta-galactosidase
893	32	39.0	611	2	A83926	hypothetical prote	966	32	39.0	1024	2	E90678	beta-D-galactosida
894	32	39.0	626	2	A82771	hypothetical prote	967	32	39.0	1024	2	A85529	beta-D-galactosida
895	32	39.0	626	2	A42891	beta-galactosidase	968	32	39.0	1025	1	JC1266	beta-galactosidase
896	32	39.0	628	2	A51780	NADH dehydrogenase	969	32	39.0	1034	2	T30551	beta-galactosidase
897	32	39.0	628	2	AF1404	NADH dehydrogenase	970	32	39.0	1034	2	T30574	beta-galactosidase
898	32	39.0	642	2	E70683	probable helix-tur	971	32	39.0	1034	2	A24925	beta-galactosidase
899	32	39.0	645	2	C84999	ATP-dependent DNA	972	32	39.0	1042	1	GBEC	beta-galactosidase
900	32	39.0	660	2	H82281	conserved hypotet	973	32	39.0	1042	2	E85968	evolved beta-D-gal
901	32	39.0	662	2	A37226	glucose transport	974	32	39.0	1042	2	F91123	evolved beta-D-gal
902	32	39.0	664	2	A33545	Na+/Glucose cotran	975	32	39.0	1048	2	T15045	ras GTPase-activat
903	32	39.0	664	2	C71106	hypothetical prote	976	32	39.0	1051	2	T48933	WD repeat domain p
904	32	39.0	671	2	S61099	leukotriene-A4 hyd	977	32	39.0	1060	2	A10201	beta-galactosidase
905	32	39.0	675	1	FVFFB	brown protein - fr	978	32	39.0	1067	2	D96545	probable DNA polym

979 32 39.0 1080 2 A35088 phycobilisome link
980 32 39.0 1085 2 S55352 IFH1 protein - yea
981 32 39.0 1087 2 F72283 beta-galactosidase
982 32 39.0 1093 2 T51503 valine-tRNA ligase
983 32 39.0 1102 2 S44772 C29E4.4 protein -
984 32 39.0 1127 2 S47445 MDML protein - yea
985 32 39.0 1137 2 D89610 ras GTPase-activat
986 32 39.0 1140 1 T38908 UV-damaged DNA-bin
987 32 39.0 1140 1 S38777 UV-damaged DNA-bin
988 32 39.0 1140 2 JC7152 probable ATP-depen
989 32 39.0 1152 2 H88533 probable chemotaxi
990 32 39.0 1161 2 H95903 ras GTPase-activat
991 32 39.0 1170 2 T19042 hypothetical prote
992 32 39.0 1185 2 T46428 phosphatidylinosit
993 32 39.0 1188 2 JC4889 SH2-containing ino
994 32 39.0 1189 2 JC6118 ras GTPase-activat
995 32 39.0 1207 2 T19041 adenosine deaminas
996 32 39.0 1226 1 S65593 PiliY protein homo
997 32 39.0 1230 2 F82857 hypothetical prote
998 32 39.0 1231 2 T24415 adenylylate cyclase
999 32 39.0 1248 2 A53588 type VIII adenylyl
1000 32 39.0 1251 2 S48687

ALIGNMENTS

RESULT 1

A32576
beta-glucuronidase (EC 3.2.1.31) allele B precursor - mouse
N;Alternate names: beta-D-glucuronoside glucuronohydrolase
C;Species: Mus musculus (house mouse)
C;Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004
C;Accession: A32576; B32576; I49692; A28954; A29977; A35798
R;Wawrzyniak, C.J.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E.
Mol. Cell. Biol. 9, 4074-4078, 1989
A;Title: DNA determinants of structural and regulatory variation within the murine beta-
A;Reference number: A32576; MUID:89384641; PMID:2779578
A;Accession: A32576
A;Molecule type: mRNA
A;Residues: 1-86, 'I', 88-648 <WAW>
A;Cross-references: UNIPARC:UPI000016CDB0; GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:
A;Experimental source: allele H
R;Funkenstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.
Mol. Cell. Biol. 8, 1160-1168, 1988
A;Title: Genomic organization and sequence of the Gus-s-a allele of the murine beta-gluc
A;Reference number: 149692; MUID:88216590; PMID:2835664
A;Accession: 149692
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-232, 'T', 234-264, 'D', 266-319, 'V', 321-427, 'K', 429-615, 'L', 617-648 <RES>
A;Cross-references: UNIPARC:UPI000016CDB5; GB:M19279; NID:g193524; PIDN:AAA37697.1; PID:
A;Experimental source: allele A
R;D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.
Biochemistry 27, 7131-7140, 1988
A;Title: Complete sequence and organization of the murine beta-glucuronidase gene.
A;Reference number: A28954; MUID:89062453; PMID:3196706
A;Accession: A28954
A;Molecule type: DNA
A;Residues: 1-264, 'D', 266-319, 'V', 321-648 <DAM>
A;Cross-references: UNIPARC:UPI000027ABB; GB:J02836; GB:J03035; GB:M20204; NID:g193716;
R;Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.
Genomics 2, 215-219, 1988
A;Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its dedu
A;Reference number: A29977; MUID:88284700; PMID:3397060
A;Accession: A29977
A;Molecule type: mRNA
A;Residues: 1-264, 'D', 266-319, 'V', 321-648 <GAL>

A;Cross-references: UNIPARC:UPI0000027ABB; GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:
J;Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.
J. Biol. Chem. 265, 14732-14735, 1990
A;Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in con
the serpin superfamily.
A;Reference number: A35798; MUID:90368633; PMID:2394691
A;Accession: A35798
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 593-648 <LIA>
A;Cross-references: UNIPARC:UPI0000175B40
A;Note: the location of the propeptide cleavage site was not demonstrated directly but re
C;Comment: in some tissues, a portion of this enzyme is retained in the endoplasmic retic
syn.
C;Genetics:
A;Gene: Gus
A;Map position: 5
A;Introns: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1
C;Superfamily: beta-glucuronidase
C;Keywords: glycosidase; hydrolase; lysosome
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-648/Product: beta-glucuronidase, ER-retained form #status predicted <ERMT>
F;23-633/Product: beta-glucuronidase, lysosomal (default) form #status predicted <LMAT>
F;634-648/Domain: carboxyl-terminal propeptide #status predicted <CPRO>

Query Match 54.9%; Score 45; DB 2; Length 648;
Best Local Similarity 57.1%; Pred. No. 7.5;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
| : |||: ||| |
DB 442 WSVANEPSSALKPA 455

RESULT 2

T49366
myocyte-specific enhancer factor 2d related protein [imported] - Neurospora crassa
N;Alternate names: protein B1D1.200
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49366
R;Schulte, U.; Aign, V.; Hohenel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49366
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-624 <SCH>
A;Cross-references: UNIPROT:Q9PGA0; UNIPARC:UPI000006A730; EMBL:ALJ355927; GSPDB:GN00116;
A;Experimental source: BAC clone B1D1; strain OR74A
C;Genetics:
A;Gene: NCSP:B1D1.200
A;Map position: 6
A;Introns: 11/1; 148/1; 167/2; 244/1

Query Match 52.4%; Score 43; DB 2; Length 624;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPN 9
: |||: ||| |
DB 238 TWQLGNSPN 246

RESULT 3

CCR7P
cytochrome c2 - Rhodopseudomonas palustris (strain 2.1.37)
C;Species: Rhodopseudomonas palustris
C;Date: 22-May-1981 #sequence_revision 22-May-1981 #text_change 31-Dec-2004
C;Accession: A00083
R;Ambler, R.P.; Daniel, M.; Hermoso, J.; Meyer, T.E.; Bartsch, R.G.; Kamen, M.D.
Nature 278, 659-660, 1979
A;Title: Cytochrome c-2 sequence variation among the recognised species of purple nonsulf

A;Reference number: A00086; MUID:79199667; PMID:221822

A;Accession: A00083

A;Molecule type: protein

A;Residues: 1-114 <AMB>

A;Cross-references: UNIPROT:P00091; UNIPARC:UPI0000128855

A;Experimental source: ATCC 17007

C;Superfamily: cytochrome c/cytochrome c2; cytochrome c homology

C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis

F;4-110/Domain: cytochrome c homology <CYC>

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F;13,16/Binding site: heme (Cys) (covalent) #status predicted

F;17,93/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 51.2%; Score 42; DB 1; Length 114;

Best Local Similarity 77.8%; Pred. No. 3.6;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 6 NEPNSFLKK 14

DB 68 NDPNAFLKK 76

RESULT 4

S48273

probable transcription factor YBR108w - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein YER0901

C;Species: *Saccharomyces cerevisiae*

C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004

C;Accession: S48273; S45976; S44688

R;Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A;Reference number: S48255; MUID:95208357; PMID:7900426

A;Accession: S48273

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-848 <MAN>

A;Cross-references: UNIPROT:P38266; UNIPARC:UPI000013A3E4; EMBL:X78993; NID:g476045; PID

R;Feldmann, H.; Mannhaupt, G.; Schwarzlöse, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S45927

A;Accession: S45976

A;Molecule type: DNA

A;Residues: 1-848 <FE2>

A;Cross-references: UNIPARC:UPI000013A3E4; EMBL:Z35977; NID:9536378; PID:g536379; MIPS:Y

C;Genetics:

A;Cross-references: SGD:S0000312

A;Map position: 2R

Query Match

Best Local Similarity 51.2%; Score 42; DB 2; Length 848;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LGNEPNSFLKK 14

DB 715 LGDAPNSFIRK 725

RESULT 5

T08469

endo-1,4-beta-xylanase (EC 3.2.1.8) - *Dictyoglomus thermophilum*

C;Species: *Dictyoglomus thermophilum*

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004

C;Accession: T08469

R;Gibbs, M.D.; Reeves, R.A.; Bergquist, P.L.

Appl. Environ. Microbiol. 61, 4403-4408, 1995

A;Title: Cloning, sequencing, and expression of a xylanase gene from the extreme thermop

A;Reference number: Z16432; MUID:96086022; PMID:8534104

A;Accession: T08469

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-352 <GTB>

A;Cross-references: UNIPROT:Q12603; UNIPARC:UPI000013907F; EMBL:L39866; NID:g973982; PID

A;Experimental source: strain Rt46B.1

C;Genetics:

A;Note: xyna

C;Function:

A;Description: hydrolyzes xylan to xylotriose and xylobiose but could not hydrolyze xylol

C;Superfamily: xylanase; Streptomyces endo-1,4-beta-xylanase A homology

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;60-352/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 50.6%; Score 41.5; DB 2; Length 352;

Best Local Similarity 40.0%; Pred. No. 16;

Matches 8; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

QY 1 SWELGNE-----PNSFLKKA 15

DB 149 AWDVVNEALSNDPNEFLRRA 168

RESULT 6

B83007

conserved hypothetical protein PA5115 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C;Accession: B83007

R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathog

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83007

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-194 <STO>

A;Cross-references: UNIPROT:Q9HU69; UNIPARC:UPI000000CSF0C; GB:AE004924; GB:AE004091; NID

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA5115

Query Match

Best Local Similarity 50.0%; Score 41; DB 2; Length 194;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 GNEPNSFLKK 14

DB 14 GTEPNAFLKE 23

RESULT 7

T19330

hypothetical protein C16C10.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T19330

R;Lloyd, C.

submitted to the EMBL Data Library, November 1994

A;Reference number: Z19108

A;Accession: T19330

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-274 <WIL>

A;Cross-references: UNIPROT:Q09465; UNIPARC:UPI0000060EFC; EMBL:Z46787; PIDN:CAA86747.1;

A;Experimental source: clone C16C10

C;Genetics:

A;Gene: CESP:C16C10.9

A;Map position: 3

A;Introns: 98/3; 117/3; 143/3; 170/2; 193/3; 222/3; 248/3

C;Superfamily: *Caenorhabditis elegans* hypothetical protein C16C10.9

Query Match 50.0%; Score 41; DB 2; Length 274;

Best Local Similarity 53.8%; Pred. No. 15;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKK 14
:|:|:|:|:|:|
Db 173 FEMNGDPFFVK 185

RESULT 8

B84900
hypothetical protein At2g46220 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: B84900
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84900
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-550 <STO>
A:Cross-references: UNIPROT:O82347; UNIPARC:UPI000017A71B; GB:AE002093; NID:g3702327; PI
C:Genetics:
A:Gene: At2g46220
A:Map position: 2

Query Match 50.0%; Score 41; DB 2; Length 550;
Best Local Similarity 61.5%; Pred. No. 33;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLK 13
||:|:|:|:|:|
Db 434 SWHLGSETKTLK 446

RESULT 9

A25047
beta-glucuronidase (BC 3.2.1.31) precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1998 #sequence_revision 30-Jun-1998 #text_change 09-Jul-2004
C:Accession: A25047; S00345
R;Nishimura, Y.; Rosenfeld, M.G.; Kreibich, G.; Gubler, U.; Sabatini, D.D.; Adesnik, M.;
Proc. Natl. Acad. Sci. U.S.A. 83, 7292-7296, 1986
A:Title: Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro
A:Reference number: A25047; MUID:87016933; PMID:3463967
A:Accession: A25047
A:Molecule type: mRNA
A:Residues: 1-648 <NIS>
A:Cross-references: UNIPROT:P06760; UNIPARC:UPI0000126911; GB:M13962; NID:g204329; PIDN:
A:Experimental source: female preputial gland
R;Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.
Biochem. J. 250, 547-555, 1988
A:Title: Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and expression
A:Reference number: S00345; MUID:88183378; PMID:3355537
A:Accession: S00345
A:Molecule type: mRNA
A:Residues: 'E', 15-20, 'L', 22-486, 'L', 488-648 <POW>
A:Cross-references: UNIPARC:UPI0000017097A; EMBL:F00717; NID:g56270; PIDN:CAA68705.1; PID
C:Superfamily: beta-glucuronidase
C:Keywords: glycosidase; hydrolase
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-648/Product: beta-glucuronidase #status predicted <MAT>

Query Match 50.0%; Score 41; DB 2; Length 648;
Best Local Similarity 57.1%; Pred. No. 40;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
|:|:|:|:|:|
Db 442 WSVANEPVSSLKPA 455

RESULT 10

A:Status: preliminary

C36955

flagellar biosynthesis-specific protein fliQ - Yersinia pestis plasmid pCD1
C:Species: Yersinia pestis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: C36955
R;Fields, K.A.; Plano, G.V.; Straley, S.C.

J. Bacteriol. 176, 569-579, 1994

A:Title: A low-Ca(2+) response (LCR) secretion (ysc) locus lies within the lcrB region of

A:Reference number: A36955; MUID:94131934; PMID:8300512

A:Accession: C36955

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <PIE>

A:Cross-references: UNIPROT:P40298; UNIPARC:UPI000016E1D4; GB:L22495; NID:g2465655; PIDN:

C:Genetics:

A:Gene: yscS

A:Genome: plasmid

C:Superfamily: flagellar biosynthesis-specific protein

Query Match 49.4%; Score 40.5; DB 1; Length 79;
Best Local Similarity 64.3%; Pred. No. 4.4;
Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 SWELGNEPNSFLKK 14
||:|:|:|:|:|
Db 67 SW-LGNELHSFVQK 79

RESULT 11

B97391

hypothetical protein AGR_C_459 [imported] - Agrobacterium tumefaciens (strain C58, Cereor

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004

C:Accession: B97391

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Oucollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: B97391

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-117 <KUR>

A:Cross-references: UNIPROT:Q8UIM6; UNIPARC:UPI00000D175E; GB:AE007869; PIDN:AAK86083.1;

C:Genetics:

A:Gene: AGR_C_459

A:Map position: circular chromosome

Query Match 48.8%; Score 40; DB 2; Length 117;
Best Local Similarity 52.6%; Pred. No. 8.4;
Matches 10; Conservative 1; Mismatches 4; Indels 4; Gaps 1;

QY 1 SWEL-----GNENPSFLKKA 15
||:|:|:|:|:|
Db 47 SWQLNRNGAELNGFLKNA 65

RESULT 12

AC2609

hypothetical protein Atu0267 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

C:Accession: AC2609

R;Wood, D.W.; Stutbal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,

A.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AC2609

A>Status: preliminary

A:Molecule type:	DNA
A:Residues:	1-117 <KUR>
A:CROSS-references:	UNIPROT:Q8UIM6; UNIPARC:UPI00000D175E; GB:AE008688; PIDN:AAL41289.1;
A:Experimental source:	strain C58 (Dupont)
C:Genetics:	
A:Gene:	Atu0267
A:Map position:	circular chromosome
Query Match	48.8%; Score 40; DB 2; Length 117;
Best Local Similarity	52.6%; Pred. No. 8.4;
Matches	10; Conservative 1; Mismatches 4; Indels 4; Gaps 1;
QY	1 SWEL-----GNPNFLKKA 15 :
DB	47 SWQLNRNGAELNGFLKNA 65
RESULT 13	
T32412	
hypothetical protein W09B6.5 - Caenorhabditis elegans	
C:Species:	Caenorhabditis elegans
C>Date:	29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession:	T32412
R:Goela, D.; Maggi, L.; Andrews, S.	
submitted to the EMBL Data Library, September 1997	
A:Description:	The sequence of C. elegans cosmid W09B6.
A:Reference number:	Z21162
A:Accession:	T32412
A>Status:	preliminary; translated from GB/EMBL/DDBJ
A:Molecule type:	DNA
A:Residues:	1-270 <GOE>
A:CROSS-references:	UNIPROT:Q9GZ14; UNIPARC:UPI0000080DB9; EMBL:AF025469; PIDN:AAB71052.
A:Experimental source:	strain Bristol N2; clone W09B6
C:Genetics:	
A:Gene:	CESP:W09B6.5
A:Map position:	2
A:Introns:	43/3; 85/3; 141/2; 192/1; 264/1
C:Superfamily:	Caenorhabditis elegans hypothetical protein W09B6.5
Query Match	48.8%; Score 40; DB 2; Length 270;
Best Local Similarity	47.1%; Pred. No. 22;
Matches	8; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
QY	1 SWELGNE--PNSFLKKA 15 : : :
DB	59 NWTLGTDFVPNFFRKA 75 : : :
RESULT 14	
S54263	
rep A protein - Bacteroides fragilis	
C:Species:	Bacteroides fragilis
C>Date:	08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C:Accession:	S54263
R:Haggoud, A.; Trinh, S.; Mohieddine, M.; Reysset, G.	
submitted to the EMBL Data Library, April 1995	
A:Description:	Genetic analysis of the minimal replicon of plasmid pIP417 and comparison
A:Reference number:	S54263
A:Accession:	S54263
A>Status:	preliminary
A:Molecule type:	DNA
A:Residues:	1-331 <HAG>
A:CROSS-references:	UNIPROT:Q45152; UNIPARC:UPI00000B0F0D; EMBL:X86702; NID:g804960; PID
Query Match	48.8%; Score 40; DB 2; Length 331;
Best Local Similarity	46.7%; Pred. No. 28;
Matches	7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY	1 SWELGNPNFLKKA 15 : :
DB	252 SWSLSSEVRSYLKNS 266

RESULT 15

B27539
variant surface glycoprotein YNat 1.3 precursor - Trypanosoma congolense
C/Species: Trypanosoma congolense
C/Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jul-2004
C/Accession: B27539
R/Strickler, J.E.; Binder, D.A.; L'Italien, J.J.; Shimamoto, G.T.; Wait, S.W.; Dalheim, I.
Biochemistry 26, 796-805, 1987
A/Title: Trypanosoma congolense: structure and molecular organization of the surface glycoprotein YNat 1.3 precursor
A/Reference number: A90525; MUID:87185370; PMID:3567147
A/Accession: B27539
A/Molecule type: mRNA
A/Residues: 1-413 <STR>
A/Cross-references: UNIPROT:P20949; UNIPARC:UPI0000138E48; GB:M15113; NID:g162439; PID:g162439
C/Keywords: glycoprotein
F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-41/Product: variant surface glycoprotein YNat 1.3 #status predicted <MAT>

Query Match 48.8%; Score 40; DB 2; Length 413;
Best Local Similarity 53.8%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WELGNEPNSFLK 14
||| ||| :||
Db 60 WALGENTSRKK 72

RESULT 16

A84709
cinamate-4-hydroxylase [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C/Accession: A84709
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Nishizawa, D.; Niernier, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: A84709
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-505 <STO>
A/Cross-references: UNIPROT:P92994; UNIPARC:UPI0000000F6E; GB:AE002093; NID:g1946370; PID:g1946370
C/Genetics:
A/Gene: At2g30490
A/Map position: 2
C/Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C/Keywords: heme; iron; metalloprotein
F/447/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 48.8%; Score 40; DB 2; Length 505;
Best Local Similarity 53.8%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLK 13
:||| ||| :
Db 400 AWLANNPNWSWK 412

RESULT 17

F81191
hypothetical protein NMB0511 [imported] - Neisseria meningitidis (strain MC58 serogroup F)
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: F81191
R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, J.
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A/Reference number: A81000; MUID:2015755; PMID:10710307

```
A;Accession: F81191
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <TET>
A;Cross-references: UNIPROT:Q9K0R9; UNIPARC:UPI0000032FA8; GB:AE002098; NID:
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0511

Query Match      48.8%; Score 40; DB 2; Length 640;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   ||| ||| |||
Db 535 WVLNVSPNDYLKEA 548

RESULT 18
A26581
beta-glucuronidase (EC 3.2.1.31) precursor - human
C;Species: Homo sapiens (man)
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C;Accession: A26581; A40337; A24983; A36538
R;Oshima, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly,
Proc. Natl. Acad. Sci. U.S.A. 84, 685-689, 1987
A;Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.
A;Reference number: A26581; MUID:87118233; PMID:3468507
A;Accession: A26581
A;Molecule type: mRNA
A;Residues: 1-651 <OSH>
A;Cross-references: UNIPROT:P08236; UNIPARC:UPI000003BC43; GB:M15182; NID:g183232; PIDN:
A;Experimental source: placenta
R;Shipley, J.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly
Genomics 10, 1009-1018, 1991
A;Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.
A;Reference number: A40337; MUID:92009900; PMID:1916806
A;Accession: A40337
A;Molecule type: DNA
A;Residues: 1-70 <SHI>
A;Cross-references: UNIPARC:UPI000016AA04; GB:M65002; NID:g183706; PIDN:AAA52622.1; PID:
R;Guise, K.S.; Korneluk, R.G.; Waye, J.; Lamhonwah, A.M.; Quan, F.; Palmer, R.; Ganschow
Gene 34, 105-110, 1985
A;Reference number: A24983; MUID:85232043; PMID:3924735
A;Accession: A24983
A;Molecule type: mRNA
A;Residues: 520-585 <GUI>
A;Cross-references: UNIPARC:UPI000016AA03; GB:M10618; NID:g183704; PIDN:AAA52621.1; PID:
R;Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeda, Y.; Yamada, S.; Yamada, Y.; Sasaki, T.;
Am J. Hum. Genet. 48, 89-96, 1991
A;Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular het
A;Reference number: A36538; MUID:91090114; PMID:1702266
A;Accession: A36538
A;Molecule type: mRNA
A;Residues: 378-385 616-621, 643-651 <TOM>
A;Cross-references: UNIPARC:UPI0000175B41
C;Genetics:
A;Gene: GDB:GUSB
A;Cross-references: GDB:120025; OMIM:253220
A;Map position: 7q22-7q22
C;Superfamily: beta-glucuronidase
C;Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-651/Product: beta-glucuronidase, placental #status predicted <MAT>

Query Match      48.8%; Score 40; DB 2; Length 651;
Best Local Similarity 50.0%; Pred. No. 60;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   ||| ||| |||
Db 446 WSVANEPASHLESA 459
```

RESULT 19

```
A35269
translation initiation factor IF-2 - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 17-Aug-1990 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: A35269; B35269; S31994; G69644
R;Shazand, K.; Tucker, J.; Chiang, R.; Stansmore, K.; Sperling-Petersen, H.U.; Grunberg-N
J. Bacteriol. 172, 2675-2687, 1990
A;Title: Isolation and molecular genetic characterization of the Bacillus subtilis gene
A;Reference number: A35269; MUID:90236932; PMID:2110148
A;Accession: A35269
A;Molecule type: DNA
A;Residues: 1-716 <SHA1>
A;Cross-references: UNIPROT:P17889; UNIPARC:UPI000012D2D0; GB:M34836; NID:g143358; PIDN:
A;Accession: B35269
A;Molecule type: DNA
A;Residues: 94-716 <SHA2>
A;Cross-references: UNIPARC:UPI0000174801
R;Shazand, K.
Submitted to the EMBL Data Library, November 1992
A;Reference number: S31990
A;Accession: S31994
A;Molecule type: DNA
A;Residues: 1-716 <SHA3>
A;Cross-references: UNIPARC:UPI000012D2D0; EMBL:Z18631; NID:g49314; PIDN:CAA79234.1; PID:
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69644
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-55, A', 57-716 <KUN>
A;Cross-references: UNIPARC:UPI0000060405; GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAF
A;Experimental source: strain 168
C;Comment: IF-2, one of the essential components for the initiation of protein synthesis
mal subunits. It is also involved in the hydrolysis of GTP during the formation of the 70
C;Comment: Alternative initiation codons in the same reading frame do appear to be utiliz
C;Genetics:
A;Gene: infB
C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog
C;Keywords: alternative initiators; GTP binding; nucleotide binding; P-loop; protein bios
F;1-716/Product: translation initiation factor IF-2-alpha #status predicted <ALP>
F;94-716/Product: translation initiation factor IF-2-beta #status predicted <BET>
F;220-329/Domain: translation elongation factor Tu homology <ETU>
F;226-233/Region: nucleotide-binding motif A (P-loop)
F;326-329/Region: GTP-binding NKXD motif
F;362-364/Region: GTP-binding SAK/L motif
F;232,233,253,326,327,329,362/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #
Query Match      48.8%; Score 40; DB 1; Length 716;
Best Local Similarity 58.3%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELGNEPNSFLKK 14
   ||| ||| |||
Db 151 ELGKEPSFLIKK 162

RESULT 20
```

D82704
conserved hypothetical protein XF1252 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: D82704
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: D82704
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1641 <STM>
A:Cross-references: UNIPROT:Q9PDX7; UNIPARC:UPI000013A4EB; GB:AE003959; NID:AE003959
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1252

Query Match 48.8%; Score 40; DB 2; Length 1641;
Best Local Similarity 70.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SWELGNPNPS 10
||||| |:
Db 212 SWELSNQNS 221

RESULT 21
S48138
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) chain VIII - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J0526; protein YJL166w; ubiquinol-cytochrome-c reductase 11K
C:Species: Saccharomyces cerevisiae
C>Date: 30-Sep-1991 #sequence_revision 11-Nov-1994 #text_change 09-Jul-2004
C:Accession: S48138; S05860; S48137; S56949
R:Koerkamp, M.G.
submitted to the EMBL Data Library, June 1993
A:Reference number: S48138
A:Accession: S48138
A:Molecule type: DNA
A:Residues: 1-94 <KOE>
A:Cross-references: UNIPROT:P08525; UNIPARC:UPI000004F931; EMBL:X05550; NID:g312353; PID E:Maarse, A.C.; Grivell, L.A.
Eur. J. Biochem. 165, 419-425, 1987
A:Title: Nucleotide sequence of the gene encoding the 11-kDa subunit of the ubiquinol-c A:Reference number: S05860; MUID:87246620; PMID:3036507
A:Accession: S05860
A:Molecule type: DNA
A:Residues: 1-61, 'LREFIGGRTVTSIMNFCTAKLVKSKELMENIV' <MAA1>
A:Cross-references: UNIPARC:UPI0000179F9B; EMBL:X05550
A:Accession: S48137
A:Molecule type: protein
A:Residues: 2-11 <MAA2>
A:Cross-references: UNIPARC:UPI0000179F9C
R:Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56937
A:Accession: S56949
A:Molecule type: DNA

A:Residues: 1-94 <OBE>
A:Cross-references: UNIPARC:UPI000004F931; EMBL:Z49441; NID:g1008355; PID:g1008356; MIPS C:Genetics:
A:Gene: SGD:OCR8
A:Cross-references: SGD:S0003702; MIPS:YJL166w
A:Map position: 10L
A:Genome: nuclear
C:Function:
A:Description: oxidoreductase
A:Pathway: oxidative phosphorylation; respiratory chain
C:Superfamily: Schizosaccharomyces pombe ubiquinol-cytochrome-c reductase chain VIII
C:Keywords: membrane-associated complex; mitochondrion; oxidative phosphorylation; oxid F:2-94/Product: ubiquinol-cytochrome-c reductase chain VIII #status experimental <MAT>

Query Match 47.6%; Score 39; DB 2; Length 94;
Best Local Similarity 63.8%; Pred. No. 9.9;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WELGNPNNSFL 12
|: ||| |:
Db 70 WKNGNEYNEFL 80

RESULT 22
T22536
hypothetical protein F53B2.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22536
R:Smyle, R.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19577
A:Accession: T22536
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-263 <WIL>
A:Cross-references: UNIPROT:Q020705; UNIPARC:UPI0000061070; EMBL:Z73908; PIDN:CAA98129.1; C:Experimental source: clone F53B2
C:Genetics:
A:Gene: CESP:F53B2.4
A:Map position: 4
A:Introns: 15/3; 85/3; 138/3; 204/3
A:Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b

Query Match 47.6%; Score 39; DB 2; Length 263;
Best Local Similarity 53.8%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNNSFLK 13
||| |:
Db 37 SWCASQQPNHFLK 49

RESULT 23
A84326
hypothetical protein Vngl1740c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84326
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: A84326
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <STO>
A:Cross-references: UNIPROT:Q9HP97; UNIPARC:UPI000006396C; GB:AE004437; NID:g10581200; C:Genetics:
A:Gene: VNG1740C

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: H94619
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-481 <STO>
 A:Cross-references: UNIPROT:Q64814; UNIPARC:UPI000000C6D0; GB:AE002093; NID:g3169178; PID:10617197
 C:Genetics:
 A:Gene: At2g23050
 A:Map position: 2
 C:Superfamily: Arabidopsis hypothetical protein F19F18.80

Query Match 47.6%; Score 39; DB 2; Length 481;
 Best Local Similarity 58.3%; Pred. No. 65;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 ELGNPNFLKK 14

DB 5 KLGKPDSEFLSK 16

RESULT 29

S67621 hypothetical protein YDL085w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein D2436
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
 C:Accession: S67621

R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67608
 A:Accession: S67621
 A:Molecule type: DNA
 A:Residues: 1-545 <WAM>
 A:Cross-references: UNIPROT:Q07500; UNIPARC:UPI000006C394; EMBL:Z74133; NID:g1431109; PID:10617197
 C:Genetics:
 A:Gene: MIPS:YDL085w
 A:Cross-references: SGD:S0002243
 A:Map position: 4L
 C:Superfamily: NADH dehydrogenase

Query Match 47.6%; Score 39; DB 2; Length 545;
 Best Local Similarity 58.3%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNPNFLK 13

DB 364 WATGNPDIFSK 375

RESULT 30

H95922 hypothetical membrane-anchored protein [imported] - *Sinorhizobium meliloti* (strain 1021)
 C:Species: *Sinorhizobium meliloti*
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: H95922

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: H95922

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-554 <KUR>
 A:Cross-references: UNIPROT:Q92VQ1; UNIPARC:UPI00000CB5E8; GB:AL591985; PIDN:CAC49048.1;
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation
 C:Genetics:
 A:Gene: SWB21069
 A:Genome: plasmid

Query Match 47.6%; Score 39; DB 2; Length 554;
 Best Local Similarity 60.0%; Pred. No. 76;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNPNFSF 11

DB 205 WELGNEMENY 214

RESULT 31

AD2408 hypothetical protein all4820 [imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: *Nostoc* sp. PCC 7120
 A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
 C:Accession: AD2408

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shimoto, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2408
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-628 <KUR>
 A:Cross-references: UNIPROT:O8YV6; UNIPARC:UPI000000C70; GB:BA000019; PIDN:BAB76519.1;
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4820

Query Match 47.6%; Score 39; DB 2; Length 628;
 Best Local Similarity 85.7%; Pred. No. 88;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNP 8

DB 419 WELGNRP 425

RESULT 32

AD3057 glycoen debranching enzyme glgX [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C:Species: *Agrobacterium tumefaciens*
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 31-Dec-2004
 C:Accession: AD3057
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; L.
 erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 sser, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AD3057
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-651 <KUR>
 A:Cross-references: UNIPROT:Q8U8L7; UNIPARC:UPI00000D242A; GB:AE008689; PIDN:AAL44874.1;
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: glgX
 A:Map position: linear chromosome
 C:Superfamily: isoamylase-type debranching enzyme

```

Query Match          47.6%; Score 39; DB 2; Length 651;
Best Local Similarity 58.3%; Pred. No. 92;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
   :||| | |||
Db 371 YQLGNFPESFLE 382

RESULT 33
B98229
glycogen debranching enzyme (AJ291603) [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 31-Dec-2004
C;Accession: B98229
A;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B98229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-651 <KUR>
A;Cross-references: UNIPROT:Q8U8L7; UNIPARC:UPI00000D242A; GB:AE007870; PIDN:AAK89356.1;
C;Genetics:
A;Gene: AGR L 1566
A;Map position: linear chromosome
C;Superfamily: isoamylase-type debranching enzyme

Query Match          47.6%; Score 39; DB 2; Length 651;
Best Local Similarity 58.3%; Pred. No. 92;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
   :||| | |||
Db 371 YQLGNFPESFLE 382

RESULT 34
I51684
epithelial sodium channel, gamma subunit - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I51684
R;Puoti, A.; May, A.; Canessa, C.M.; Horisberger, J.
Am. J. Physiol. 269, 188-197, 1995
A;Title: The highly selective, low conductance epithelial sodium channel of Xenopus lae
A;Reference number: I51682
A;Accession: I51684
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-660 <PUO>
A;Cross-references: UNIPROT:PS1171; UNIPARC:UPI0000135621; EMBL:U25342; NID:g886045; PID
C;Genetics:
A;Gene: GammaENaC
C;Superfamily: human amiloride-sensitive sodium channel protein; fibronectin type I repe
F;375-410/Domain: fibronectin type I repeat homology <IFR>

Query Match          47.6%; Score 39; DB 2; Length 660;
Best Local Similarity 57.1%; Pred. No. 93;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLK 14
   ||||| : | |
Db 488 SWELGEXLKNLTK 501

RESULT 35
D72278
endo-1,4-beta-mannosidase - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

```

```

C;Accession: D72278
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72278
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-669 <ARN>
A;Cross-references: UNIPROT:Q9X0V4; UNIPARC:UPI00000D38F2; GB:AE001779; GB:AE000512; NID:
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1227

Query Match          47.6%; Score 39; DB 2; Length 669;
Best Local Similarity 75.0%; Pred. No. 94;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEP 8
   :||| | |||
Db 193 AWELANEP 200

RESULT 36
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Bacillus sp.
C;Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S29043; PC4404
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A;Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A;Reference number: S29043; MUID:91037937; PMID:2230718
A;Accession: S29043
A;Molecule type: DNA
A;Residues: 1-941 <OZA>
A;Cross-references: UNIPROT:P19424; UNIPARC:UPI000012BDFB; EMBL:M27420; NID:g142664; PIDN
R;Shirai, T.; Yamane, T.; Hida, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito,
J. Biochem. 122, 683-685, 1997
A;Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A;Reference number: PC4404; MUID:98060488; PMID:9399567
A;Accession: PC4404
A;Molecule type: protein
A;Residues: 228-584 <SHI>
A;Cross-references: UNIPARC:UPI00001791AD
A;Experimental source: strain KSM-635
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A;Pathway: cellulose degradation
C;Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermot
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;41-95/Domain: S-layer repeat homology <SLR1>
F;101-153/Domain: S-layer repeat homology <SLR2>
F;164-219/Domain: S-layer repeat homology <SLR3>
F;766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match          47.6%; Score 39; DB 2; Length 941;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNEPN 9
   ||| | |||
Db 368 WELANEPS 375

RESULT 37
E72215
oligopeptide ABC transporter, periplasmic oligopeptide-binding protein - Thermotoga marit
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

```

C;Accession: E72215
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
 C.M.
 Nature 399, 323-329, 1999
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A;Reference number: A72200; PMID:99287316; PMID:10360571
 A;Accession: E72215
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-642 <ARN>
 A;Cross-references: UNIPROT:Q9X268; UNIPARC:UPI00000D3860; GB:AE001813; GB:AE000512; NID
 A;Experimental source: strain MSB8
 C;Genetics:
 A;Gene: TM1746

Query Match 47.0%; Score 38.5; DB 2; Length 642;
 Best Local Similarity 44.4%; Pred. No. 1;le+02;
 Matches 8; Conservative 3; Mismatches 2; Indels 5; Gaps 1;
 QY 2 WE-----LGNPNPNSFLKK 14
 ||| : ||| |||
 Db 490 WEKQEVNNSPDEFLKK 507
 ||| : ||| |||

RESULT 38
 G96948
 uncharacterized small conserved protein, homolog of YUKF/YFJA B. subtilis CAC0398 [import
 C;Species: Clostridium acetobutylicum
 C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C;Accession: G96948
 R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 ; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
 A;Reference number: A96900; PMID:21359325; PMID:21359325
 A;Accession: G96948
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-100 <KUR>
 A;Cross-references: UNIPROT:Q97M02; UNIPARC:UPI00000C9B5E; GB:AE001437; PIDN:AAK78378.1;
 A;Experimental source: Clostridium acetobutylicum ATCC824
 C;Genetics:
 A;Gene: CAC0398

Query Match 46.3%; Score 38; DB 2; Length 100;
 Best Local Similarity 54.5%; Pred. No. 16;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SWELGNEPNSF 11
 ||| | | |
 Db 73 AWELNNTATPF 83
 ||| | | |

RESULT 39
 H87375
 hypothetical protein CC1020 [imported] - Caulobacter crescentus
 C;Species: Caulobacter crescentus
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C;Accession: H87375
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Koon
 N., J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of Caulobacter crescentus.
 A;Reference number: A87249; PMID:21173698; PMID:11259647
 A;Accession: H87375
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-129 <STO>
 A;Cross-references: UNIPROT:Q9A9G8; UNIPARC:UPI00000C7244; GB:AE005673; NID:gl3422312; F
 C;Genetics:
 A;Gene: CC1020

Query Match 46.3%; Score 38; DB 2; Length 129;
 Best Local Similarity 42.9%; Pred. No. 22;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKK 14
 : ||| | | |
 Db 108 AMLVGSEPKSLKR 121
 : ||| | | |

RESULT 40
 F86239
 protein F20B24.4 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C;Accession: F86239
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; PMID:21016719; PMID:11130712
 A;Accession: F86239
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-135 <STO>
 A;Cross-references: UNIPROT:Q9SGV9; UNIPARC:UPI000009D28C; GB:AE005172; NID:96573734; P
 C;Genetics:
 A;Gene: F20B24.4
 A;Map position: 1

Query Match 46.3%; Score 38; DB 2; Length 135;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 WEELGNEPNSF 11
 ||| | | |
 Db 29 WKLGDDPSRF 38
 ||| | | |

RESULT 41
 AH1586
 bacteriophage phi-105 ORF2 protein homolog lin1233 [imported] - Listeria innocua (strain
 C;Species: Listeria innocua
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AH1586
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H.
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A;Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; PMID:21537279; PMID:11679669
 A;Accession: AH1586
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-139 <GLA>
 A;Cross-references: UNIPROT:Q92CD9; UNIPARC:UPI00000CC4E3; GB:AL592022; PIDN:CAC96464.1;
 A;Experimental source: strain Clp11262
 C;Genetics:
 A;Gene: lin1233
 C;Superfamily: phage phi-105 immunity region protein 2

Query Match 46.3%; Score 38; DB 2; Length 139;
 Best Local Similarity 50.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPSFLK 13
| | | | |
Db 33 WNLANDTNGFYK 44

RESULT 42
822788
thymidylate kinase XF0580 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 31-Dec-2004
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: F82788
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <STM>
A:Cross-references: UNIPARC:UPI0000165A56; GB:AE003904; GB:AE003849; NID:g9105433; PIDN:
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tshuko, da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshuko, da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0580
C:Superfamily: thymidylate kinase

Query Match 46.3%; Score 38; DB 2; Length 217;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WELGNEPSNF 11
| | | | |
Db 146 WERGSTNHP 155

RESULT 43
AE2956
haloacid dehalogenase-like hydrolase [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 05-Oct-2004
C:Accession: AE2956
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, E.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE2956
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <KUR>
A:Cross-references: UNIPROT:Q8UAW6; UNIPARC:UPI00000D212D; GB:AE008689; PIDN:AAL44067.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3251
A:Map position: linear chromosome
C:Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 46.3%; Score 38; DB 2; Length 228;
Best Local Similarity 46.7%; Pred. No. 42;
Matches 7; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 2 WELGN--EPNSFLKK 14
| | | | |
Db 41 WAVANGIEPNAFLQ 55

RESULT 44

A98327
hypothetical protein AGR_L_3128 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 05-Oct-2004
C:Accession: A98327
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: A98327
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <KUR>
A:Cross-references: UNIPROT:Q8UAW6; UNIPARC:UPI00000D212D; GB:AE007870; PIDN:AAK90139.1;
C:Genetics:
A:Gene: AGR_L_3128
A:Map position: linear chromosome
C:Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 46.3%; Score 38; DB 2; Length 228;
Best Local Similarity 46.7%; Pred. No. 42;
Matches 7; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 2 WELGN--EPNSFLKK 14
| | | | |
Db 41 WAVANGIEPNAFLQ 55

RESULT 45

S49780
hypothetical protein YDR183w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YD9395.17
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 09-Jul-2004
C:Accession: S49780
R:Murphy, L.; Harris, D.E.
submitted to the EMBL Data Library, November 1994
A:Reference number: S49764
A:Accession: S49780
A:Molecule type: DNA
A:Residues: 1-230 <MUR>
A:Cross-references: UNIPROT:Q04004; UNIPARC:UPI000013A8BF; EMBL:Z46727; NID:gl289283; PI
C:Genetics:
A:Gene: SGD:PLP1; MIPS:YDR183w
A:Cross-references: SGD:S0002591
A:Map position: 4R

Query Match 46.3%; Score 38; DB 2; Length 230;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ELGNPNSF 11
| | | | |
Db 174 KLGNPNGF 182

RESULT 46

F70710
probable 3-oxoacyl-[acyl-carrier protein] reductase - Mycobacterium tuberculosis (strain
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Oct-2004

C;Accession: F70710
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70710
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-247 <COL>
A;Cross-references: UNIPROT:Q48930; UNIPARC:UPI000012A460; GB:Z79701; GB:AL123456; NID:9
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: fabI
C;Superfamily: short-chain dehydrogenase; short-chain alcohol dehydrogenase homology
F;16-184/Domain: short-chain alcohol dehydrogenase homology <SADH>
Query Match 46.3%; Score 38; DB 2; Length 247;
Best Local Similarity 46.7%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
||| :||| :|||
Db 144 SWGIGNQNYAASKA 158
RESULT 47
T38320
hypothetical protein SPAC23H4.08 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T38320
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: 221733
A;Accession: T38320
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-277 <BRO>
A;Cross-references: UNIPROT:O13951; UNIPARC:UPI000006AEOB; EMBL:Z98977; PIDN:CAB11662.1;
A;Experimental source: strain 972h-; cosmid C23H4
C;Genetics:
A;Gene: SPDB:SPAC23H4.08
A;Map position: 1
A;Introns: 21/1; 44/1; 148/2
Query Match 46.3%; Score 38; DB 2; Length 277;
Best Local Similarity 87.5%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ELGNEPNS 10
||||| :||| :|||
Db 23 ELGNNPNS 30
RESULT 48
T24356
hypothetical protein T02D1.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24356
R;Basham, V.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z19880
A;Accession: T24356
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-322 <WIL>
A;Cross-references: UNIPROT:O45729; UNIPARC:UPI0000061119; EMBL:Z83319; PIDN:CAB05908.1;
A;Experimental source: clone T02D1
C;Genetics:

A;Gene: CESP:T02D1.3
A;Map position: 4
A;Introns: 77/3; 147/3; 200/3; 266/3
C;Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b
Query Match 46.3%; Score 38; DB 2; Length 322;
Best Local Similarity 53.8%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLK 13
||| :||| :|||
Db 99 SWCASQKPNHFLK 111
RESULT 49
S09283
fructose-bisphosphate aldolase (EC 4.1.2.13) - Corynebacterium glutamicum
C;Species: Corynebacterium glutamicum
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: S09283
R;Von der Osten, C.H.; Barbas, C.F.; Wong, C.H.; Sinskey, A.J.
Mol. Microbiol. 3, 1625-1637, 1989
A;Title: Molecular cloning, nucleotide sequence and fine-structural analysis of the Corynebacterium glutamicum aldolase gene
A;Reference number: S09283; MUID:90136092; PMID:2615658
A;Accession: S09283
A;Molecule type: DNA
A;Residues: 1-344 <OST>
A;Cross-references: UNIPROT:P19537; UNIPARC:UPI000016EAEA; EMBL:X17313; NID:G40494; PIDN:
C;Genetics:
A;Gene: fda
C;Superfamily: fructose-bisphosphate aldolase II
C;Keywords: aldehyde-lyase; carbon-carbon lyase; gluconeogenesis; glycolysis
Query Match 46.3%; Score 38; DB 2; Length 344;
Best Local Similarity 47.1%; Pred. No. 67;
Matches 8; Conservative 4; Mismatches 1; Indels 4; Gaps 1;
QY 3 ELGN----EPNSFLKKA 15
|:|:| :|:| :|:|
Db 304 EVGNKKAYDPRSYMKA 320
RESULT 50
T04762
chitinase homolog T16H5.170 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T04762
R;Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer, submitted to the Protein Sequence Database, June 1998
A;Reference number: Z15383
A;Accession: T04762
A;Molecule type: DNA
A;Residues: 1-379 <BEV>
A;Cross-references: UNIPROT:O81862; UNIPARC:UPI000009E965; EMBL:AL024486
A;Experimental source: cultivar Columbia; BAC clone T16H5
C;Genetics:
A;Map position: 4
A;Introns: 232/1; 362/1
A;Note: T16H5.170
C;Superfamily: Streptomyces chitinase chi40
Query Match 46.3%; Score 38; DB 2; Length 379;
Best Local Similarity 46.7%; Pred. No. 75;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
||| :||| :|||
Db 347 SWHVGADDNSGLSRA 361
Search completed: June 5, 2006, 12:53:44

Job time : 32.6438 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 131.507 Seconds

(without alignments)

105.510 Million cell updates/sec

Title: US-10-645-659A-8

Perfect score: 82

Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : UniProt 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	543	1 HPSE HUMAN	Q9Y251 homo sapien
2	76	92.7	535	1 HPSE MOUSE	Q6Ygz1 mus musculus
3	76	92.7	536	1 HPSE RAT	Q71rp1 rattus norv
4	76	92.7	558	2 Q333X5 SPJJD	Q333x5 spalax juda
5	76	92.7	574	2 Q333X6 SPJJD	Q333x6 spalax juda
6	76	92.7	574	2 Q333X7 9RODE	Q333x7 spalax carm
7	76	92.7	574	2 Q333X8 9RODE	Q333x8 spalax gola
8	76	92.7	574	2 Q333X9 9RODE	Q333x9 spalax gali
9	73	89.0	523	1 HPSE CHICK	Q90yk5 gallus gall
10	73	89.0	545	1 HPSE BOVIN	Q9myv0 bos taurus
11	72	87.8	255	2 QATGCB TTING	Q49c8 tetraodon n
12	72	87.8	533	2 Q4SYFE TTING	Q49yf6 tetraodon n
13	58	70.7	592	1 HPSE2 HUMAN	Q8wwq2 homo sapien
14	58	70.7	592	2 Q2M1H5 HUMAN	Q2mlh9 homo sapien
15	58	70.7	597	2 Q4TB80 TTING	Q4tb80 tetraodon n
16	57	69.5	515	2 Q8T108 BOMMO	Q8t108 bombyx mori
17	52	63.4	651	1 BGLR CANFA	O18835 canis famil
18	51	62.2	652	2 Q4PAT7 FIG	Q4fat7 sus scrofa
19	50	61.0	651	1 BGLR FELCA	Q97524 felis silve
20	49	59.8	581	2 Q6FYV6 BARQU	Q6fyv6 bartonella
21	48	58.5	260	2 Q2TYD4 ASPOR	Q2tyd4 aspergillus
22	48	58.5	435	2 Q4Y355 BACFR	Q64y35 bacteroides
23	47	57.3	1297	2 Q31KG9 PSEHT	Q31kg9 pseudosalter
24	46	56.1	627	2 Q3CKA3 THEET	Q3cka3 thermoanaer
25	46	56.1	637	2 Q8RDH1 THETN	Q8rdh1 thermoanaer
26	46	56.1	651	2 Q3E4N3 CHLAU	Q3e4n3 chloroflexu
27	45	54.9	373	2 Q2U2I3 ASPOR	Q2u2i3 aspergillus
28	45	54.9	409	2 Q5B833 EMENI	Q5b833 aspergillus
29	45	54.9	437	2 Q99036 TRIRE	Q99036 trichoderma
30	45	54.9	490	2 Q3DZN2 CHLAU	Q3dzn2 chloroflexu
31	45	54.9	538	2 Q5CW71 CRYPV	Q5cw71 cryptospori

32	45	54.9	648	1 BGLR MOUSE	P12265 mus musculus
33	45	54.9	648	2 Q3TW82 MOUSE	Q3tw82 m osteoclas
34	45	54.9	648	2 Q6IR10 MOUSE	Q6ir10 m glucuroni
35	45	54.9	1687	2 Q4IR74 GIBZE	Q4ir74 gibberella
36	44	53.7	90	2 Q3D554 STRAG	Q3d554 streptococc
37	44	53.7	90	2 Q3DNV8 STRAG	Q3dnv8 streptococc
38	44	53.7	90	2 Q48S82 STRPM	Q48s82 streptococc
39	44	53.7	90	2 Q8DX45 STRA5	Q8dx45 streptococc
40	44	53.7	144	2 Q4G3Z5 MACMU	Q4g3z5 macaca mula
41	44	53.7	379	1 TSYL1 MOUSE	Q8852 mus musculus
42	44	53.7	379	2 Q3TKW0 MOUSE	Q3tkw0 m blastocys
43	44	53.7	379	2 Q642B1 RAT	Q642b1 rattus norv
44	44	53.7	437	2 QSTFE6 HUMAN	Qstfe6 homo sapien
45	44	53.7	438	1 TSYL1 HUMAN	Q9h0u9 homo sapien
46	44	53.7	438	2 Q6FI91 HUMAN	Q6fi91 homo sapien
47	44	53.7	438	2 Q5R5G8 PONPY	Q5r5g8 pongo pygma
48	44	53.7	438	2 Q5RA88 PONPY	Q5ra88 pongo pygma
49	44	53.7	497	2 Q5QNQ4 MOUSE	Q5gnq4 mus musculu
50	43.5	53.0	492	2 Q47BM1 DECAR	Q47bm1 dechloromon
51	43	52.4	75	2 Q4INV6 METBU	Q4inv6 methanococc
52	43	52.4	170	2 Q5LQE4 SILPO	Q5lqe4 sillicibacte
53	43	52.4	201	2 Q3EJL4 BACTI	Q3ejl4 bacillus th
54	43	52.4	310	2 Q55BQ2 D1CDI	Q55bq2 dictyosteli
55	43	52.4	321	2 Q3L5C1 BIFBR	Q3l5c1 bifidobacte
56	43	52.4	321	2 Q8G872 BIFLO	Q8g872 bifidobacte
57	43	52.4	365	2 Q4YT0_9BURK	Q4yt0 polaromonas
58	43	52.4	450	2 Q6ISB8 CABER	Q6isb8 caenorhabdi
59	43	52.4	623	2 Q7RY27 NEUCR	Q7ry27 neurospora
60	43	52.4	624	2 Q9PEA0 NEUCR	Q9pea0 neurospora
61	43	52.4	701	2 Q600U1 MYCH2	Q600u1 mycoplasma
62	43	52.4	717	2 Q4A816 MYCH7	Q4a816 mycoplasma
63	43	52.4	717	2 Q4A9Y6 MYCHJ	Q4a9y6 mycoplasma
64	43	52.4	1080	2 Q9CM43 PASMU	Q9cm43 pasteurella
65	43	52.4	2288	2 Q5AXQ5 EMENI	Q5axq5 aspergillus
66	42.5	51.8	166	2 Q4H5P7_9DEIO	Q4h5p7 deinococcus
67	42	51.2	59	2 Q3YOH0 ENTFC	Q3y0h0 enterococcu
68	42	51.2	88	2 Q7TTQ6 PROMM	Q7ttq6 prochloroco
69	42	51.2	110	2 Q87GD2 VIBPA	Q87gd2 vibrio para
70	42	51.2	127	2 Q39XW1 GEOMG	Q39xw1 geobacter m
71	42	51.2	139	1 CYC22 RHOPA	P00091 rhodopsendo
72	42	51.2	139	2 Q8GI80 RHOPA	Q8gi80 rhodopsendo
73	42	51.2	163	2 Q8RME6 RHOPA	Q8rme6 rhodopsendo
74	42	51.2	250	1 THYX CORGL	P40111 corynebacte
75	42	51.2	321	2 Q61YY6 CAEBR	Q61yy6 caenorhabdi
76	42	51.2	416	2 Q2SQ73 9GAMM	Q2sq73 hahelia che
77	42	51.2	436	2 Q421D4 DESHA	Q421d4 desulfitoba
78	42	51.2	489	2 Q4WH6 GIBZE	Q4wh6 gibberella
79	42	51.2	552	2 Q4IBY1 GIBZE	Q4ibyl gibberella
80	42	51.2	558	2 Q4CFT4 CLOPM	Q4cft4 clostridium
81	42	51.2	630	2 Q6CT20 KLULA	Q6ct20 kluyveromyc
82	42	51.2	758	2 Q5AF48 CANAL	Q5af48 candida alb
83	42	51.2	779	2 Q3FSA5_9BURK	Q3fsa5 burkholderi
84	42	51.2	779	2 Q4BIB6 BURVI	Q4bib6 burkholderi
85	42	51.2	904	2 Q87JH7 VIBPA	Q87jh7 vibrio para
86	42	51.2	947	1 YBV6 YEAST	P38266 saccharomyc
87	42	51.2	947	2 Q2SSN6 9SPHI	Q2ssn6 salinibacte
88	42	51.2	1420	2 Q5AFT3 CANAL	Q5aft3 candida alb
89	42	51.2	4760	2 Q3VV55 PROAE	Q3vv55 prosthecoc
90	41.5	50.6	118	2 Q81CK7 BACCR	Q81ck7 bacillus ce
91	41.5	50.6	352	1 XYNA DICTH	Q12603 dictyoglomu
92	41	50.0	48	2 Q5G5M1 9HIV1	Q5g5m1 human immu
93	41	50.0	81	2 Q3SDLO NEUCR	Q3sdlo neurospora
94	41	50.0	87	2 Q3HP06 9HIV1	Q3hp06 human immu
95	41	50.0	97	2 Q4A8N6 MYCH7	Q4a8n6 mycoplasma
96	41	50.0	97	2 Q4AAK5 MYCHJ	Q4aa5 mycoplasma
97	41	50.0	121	2 Q5X4L3 LEGPA	Q5x4l3 legionella
98	41	50.0	121	2 Q3UX22 MOUSE	Q3ux22 mus musculu
99	41	50.0	123	2 Q5WM03 LEGPL	Q5wm03 legionella
100	41	50.0	123	2 Q5ZUU4 LEGPH	Q5zuu4 legionella
101	41	50.0	127	2 Q8IHU0 PLAF7	Q8ihu0 plasmodium
102	41	50.0	158	2 Q6NJF3 CORDI	Q6njf3 corynebacte
103	41	50.0	163	2 Q50U98 ENTHI	Q50u98 entamoeba h
104	41	50.0	172	2 Q4HQZ1 CAMUP	Q4hqz1 campylobact

835	36.5	44.5	325	2	Q83TI4	CLODI	Q83ti4	clostridium	908	36	43.9	215	2	Q86G71	DERVA	Q86g71	dermaceutor
836	36.5	44.5	344	2	Q4MZN4	THEPA	Q4mzn4	theileria p	909	36	43.9	220	2	Q7PGQ5	ANOGA	Q7pgq5	anopheles g
837	36.5	44.5	348	2	Q3GDT7	9EIRM	Q3gdt7	syntrophomo	910	36	43.9	222	1	DSBA	ENTAM	Q9xqpl	enterobacte
838	36.5	44.5	450	1	GLMM	AGRT5	Q8u919	agrobacteri	911	36	43.9	225	1	FLIH	BUCPB	Q89a8	buchnera ap
839	36.5	44.5	450	2	Q2K4M3	RHET	Q2k4m3	rhizobium e	912	36	43.9	227	2	Q4CO19	CROWT	Q4c019	crocosphaer
840	36.5	44.5	451	1	GLMM	BRUSE	Q8yiu8	brucella me	913	36	43.9	229	2	Q6PL57	ELECO	Q6pl57	eleusine co
841	36.5	44.5	451	1	GLMM	BRUSE	Q8fz13	brucella su	914	36	43.9	229	2	Q8X5C5	ECOS7	Q8x5c5	escherichia
842	36.5	44.5	451	2	Q57BJ0	BRUAB	Q57bj0	brucella ab	915	36	43.9	233	2	Q88UP6	LACPL	Q88up6	lactobacill
843	36.5	44.5	451	2	Q2YQHR	BRUA2	Q2yqhr	brucella ab	916	36	43.9	235	2	Q96174	PLAF7	Q96174	plasmodium
844	36.5	44.5	460	2	Q5CKB0	CRYHO	Q5ckb0	cryptospori	917	36	43.9	235	2	Q2WN99	CLOBE	Q2wn99	clostridium
845	36.5	44.5	461	2	Q5CT31	CRYPV	Q5ct31	cryptospori	918	36	43.9	245	2	Q2NIZ3	9MOLU	Q2niz3	aeter yello
846	36.5	44.5	465	2	Q8DTA8	STRMU	Q8dta8	streptococc	919	36	43.9	245	2	Q6QYV5	ONYPE	Q6qyv5	onion yello
847	36.5	44.5	475	2	Q3YOV8	ENTFC	Q3yov8	enterococu	920	36	43.9	251	2	Q44JT8	CHRS1	Q44jt8	chromohalob
848	36.5	44.5	506	2	Q3SDX4	PARTF	Q3sdx4	paramecium	921	36	43.9	253	2	Q7WTF6	9ACTO	Q7wtf6	streptomyce
849	36.5	44.5	509	2	Q3SCX6	PARTF	Q3scx6	paramecium	922	36	43.9	254	2	Q742F0	MYCPA	Q742f0	mycobacteri
850	36.5	44.5	560	2	Q6AB58	PROAC	Q6ab58	propionibac	923	36	43.9	259	2	Q61YY3	CABER	Q61yy3	caenorhabdi
851	36.5	44.5	579	2	Q3N2K0	9DELT	Q3n2k0	syntrophoba	924	36	43.9	261	1	Y202	ENCUC	Y202	encephalito
852	36.5	44.5	643	2	Q5VMB3	ORYSA	Q5vmb3	oryza sativ	925	36	43.9	261	1	Q361Z1	MARHY	Q361z1	marinobacte
853	36.5	44.5	648	2	Q6NC95	RHOPA	Q6nc95	rhodopseudo	926	36	43.9	263	1	KCTD2	HUMAN	KCTD2	homo sapien
854	36.5	44.5	720	2	Q37DD6	RHOPA	Q37dd6	rhodopseudo	927	36	43.9	263	2	Q51716	BACNO	Q51716	bacteroides
855	36.5	44.5	728	2	Q6XQB2	9CAUD	Q6xqb2	enterobacte	928	36	43.9	263	2	Q4VBE7	MOUSE	Q4vbe7	mus muscucu
856	36.5	44.5	734	2	Q35DA6	9BRAD	Q35da6	bradyrhizob	929	36	43.9	264	2	Q5KRH7	CORGL	Q5krh7	corynebacte
857	36.5	44.5	734	2	Q37989	RHOPA	Q37989	rhodopseudo	930	36	43.9	265	2	Q8NZZ4	STRP8	Q8nzz4	streptococc
858	36.5	44.5	899	2	Q3ILA4	PSEHT	Q3ila4	pseudocalter	931	36	43.9	266	1	KCTD2	MOUSE	Q8ce20	mus muscucu
859	36.5	44.5	928	2	Q4CCP8	CLOTM	Q4ccp8	c clostridi	932	36	43.9	271	2	Q3C8G2	9CLOT	Q3c8g2	alkaliphilu
860	36.5	44.5	954	2	Q4B5Q4	9BURK	Q4b5q4	polaromonas	933	36	43.9	272	2	Q9HAE5	HUMAN	Q9hae5	homo sapien
861	36	43.9	35	2	Q97RG6	STRPN	Q97rg6	streptococc	934	36	43.9	272	2	Q9LYV6	ARATH	Q9lyv6	arabidopais
862	36	43.9	73	2	Q5WFV9	BACSK	Q5wfv9	bacillus cl	935	36	43.9	273	2	Q3CYV8	STRAG	Q3cyv8	streptococc
863	36	43.9	79	2	Q5WD50	BACSK	Q5wd50	bacillus cl	936	36	43.9	273	2	Q3D5M2	STRAG	Q3d5m2	streptococc
864	36	43.9	91	2	Q8GTA2	DATFE	Q8gta2	dataura fero	937	36	43.9	273	2	Q3DIW0	STRAG	Q3diw0	streptococc
865	36	43.9	93	2	Q4XAN3	PLACH	Q4xan3	plasmodium	938	36	43.9	273	2	Q3K2X4	STRAL	Q3k2x4	streptococc
866	36	43.9	112	2	Q9P117	HUMAN	Q9p117	homo sapien	939	36	43.9	273	2	Q8E1E6	STRAS	Q8e1e6	streptococc
867	36	43.9	115	2	Q49UG6	STAS1	Q49ug6	staphylococ	940	36	43.9	273	2	Q8B6W3	STRAS	Q8b6w3	streptococc
868	36	43.9	115	2	Q4LAA4	STAHJ	Q4laa4	staphylococ	941	36	43.9	277	2	Q6RVX9	BIFAN	Q6rvx9	bifidobacte
869	36	43.9	116	2	Q5A0A4	CANAL	Q5a0a4	candida alb	942	36	43.9	278	2	Q2Y6N4	NITMU	Q2y6n4	nitrosospir
870	36	43.9	120	2	Q41XR8	DESHA	Q41xr8	desulfitoba	943	36	43.9	281	2	Q9HPJ3	HALSA	Q9hpj3	halobacteri
871	36	43.9	125	2	Q9D3W9	MOUSE	Q9d3w9	mus muscucu	944	36	43.9	281	2	Q6LKN1	PHOPR	Q6lkn1	photobacter
872	36	43.9	130	2	Q2SHU9	9GAMM	Q2shu9	hahella che	945	36	43.9	283	2	Q5ICQ7	ENTHI	Q5icq7	entamoeba h
873	36	43.9	131	2	Q8TFH6	SCHPO	Q8tfh6	schizosacch	946	36	43.9	284	2	Q7V0Q4	PROMP	Q7v0q4	prochloroco
874	36	43.9	131	2	Q6IL14	DROME	Q6il14	drosophilila	947	36	43.9	286	2	Q6MRJ0	BDEBA	Q6mrj0	bdellovibri
875	36	43.9	138	2	P73331	SYNY3	P73331	synechocyst	948	36	43.9	288	1	DJLA	HAEN	DJLA	haemophilus
876	36	43.9	138	2	P73781	SYNY3	P73781	synechocyst	949	36	43.9	288	2	Q2WJ64	CLOBE	Q2wj64	clostridium
877	36	43.9	138	2	P73924	SYNY3	P73924	synechocyst	950	36	43.9	288	2	Q4QNR8	HAS18	Q4qnr8	haemophilus
878	36	43.9	139	2	Q2ISY0	RHOPA	Q2isy0	rhodopseudo	951	36	43.9	291	2	Q5L3S5	GEOKA	Q5l3s5	geobacillus
879	36	43.9	140	2	Q2P3J1	XANOR	Q2p3j1	xanthomonas	952	36	43.9	292	2	Q692L5	YEREN	Q692l5	yersinia en
880	36	43.9	144	2	Q4XUE7	PLACH	Q4xue7	plasmodium	953	36	43.9	294	2	Q39YH9	GEOMG	Q39yh9	geobacter m
881	36	43.9	144	2	Q8VUB4	9BRAD	Q8vub4	bradyrhizob	954	36	43.9	295	2	Q3DCG8	STRAG	Q3dcg8	streptococc
882	36	43.9	146	2	Q7G6E0	ORYSA	Q7g6e0	oryza sativ	955	36	43.9	295	2	Q3DQH3	STRAG	Q3dqh3	streptococc
883	36	43.9	147	2	Q34BK2	RHOPA	Q34bk2	rhodopseudo	956	36	43.9	301	2	Q4FSES	PSYAR	Q4fses	psychobact
884	36	43.9	152	1	G12	BRARE	P32002	brachydanio	957	36	43.9	302	2	Q82Q71	STRAM	Q82q71	streptomyce
885	36	43.9	154	1	SRCB	BUCAP	P32002	buchnera ap	958	36	43.9	302	2	Q3NRS4	SHEFR	Q3nrs4	shewanella
886	36	43.9	154	2	Q40JE3	HRCH	Q40je3	ehrlichia c	959	36	43.9	303	1	PECR	RAT	PECR	rattus norv
887	36	43.9	155	2	Q5CPL5	CRYPV	Q5cpl5	cryptospori	960	36	43.9	306	2	Q46HJ6	PROMT	Q46hj6	prochloroco
888	36	43.9	156	2	Q9YME3	9HIV1	Q9yme3	human immun	961	36	43.9	311	2	Q2IXS5	RHOPA	Q2ixs5	rhodopseudo
889	36	43.9	171	2	Q441Y2	SOLUS	Q441y2	solibacter	962	36	43.9	312	2	Q4PFE9	USTMA	Q4pfe9	ustiliago ma
890	36	43.9	172	2	Q7MXP8	PORGI	Q7mnp8	porphyromon	963	36	43.9	313	2	Q6BV50	DEBHA	Q6bv50	debaryomye
891	36	43.9	173	1	PLIV	SALMU	P37588	salmonella	964	36	43.9	313	2	Q5CJF2	CRYHO	Q5cjf2	cryptospori
892	36	43.9	174	2	Q37BN5	RHOPA	Q37bn5	rhodopseudo	965	36	43.9	313	2	Q5CVC3	CRYPV	Q5cvc3	cryptospori
893	36	43.9	175	2	Q2W619	MAGSA	Q2w619	magnetospir	966	36	43.9	314	2	Q9FI45	ARATH	Q9fi45	arabidopais
894	36	43.9	175	2	Q3GLB2	9GAMM	Q3glb2	psychrobact	967	36	43.9	315	2	Q4EMN8	LISMO	Q4emn8	listeria mo
895	36	43.9	177	2	Q8F1G8	ECOL6	Q8f1g8	escherichia	968	36	43.9	315	2	Q92TU3	LISIN	Q92tu3	listeria in
896	36	43.9	183	2	Q86K28	DICDI	Q86k28	dictyosteli	969	36	43.9	315	2	Q93RN0	LISMO	Q93rn0	listeria mo
897	36	43.9	183	2	Q9LNS6	ARATH	Q9lns6	arabidopais	970	36	43.9	316	2	Q4S6L2	TETNG	Q4s6l2	tetrasodon n
898	36	43.9	194	2	Q3IF16	PSEHT	Q3if16	pseudocalter	971	36	43.9	323	2	Q62050	CABEL	Q62050	caenorhabdi
899	36	43.9	201	2	Q4UDH8	THEAN	Q4udh8	theileria a	972	36	43.9	323	2	Q3YS21	EHRGJ	Q3ye21	ehrlichia c
900	36	43.9	201	2	Q9ASV1	ARATH	Q9asv1	arabidopais	973	36	43.9	324	2	Q7QH17	ANOGA	Q7qh17	anopheles g
901	36	43.9	201	2	Q3HF65	TRIER	Q3hf65	trichodesmi	974	36	43.9	324	2	Q7VXB3	BORPE	Q7vxb3	bordetella
902	36	43.9	203	2	Q97YK0	SULSO	Q97yk0	sulfolobus	975	36	43.9	325	2	Q3KZ21	BIFAN	Q3kz21	bifidobacte
903	36	43.9	207	2	Q6FU74	CANGA	Q6fu74	candida gla	976	36	43.9	326	2	Q5CU30	CRYPV	Q5cu30	cryptospori
904	36	43.9	209	2	Q7AF16	ECOS7	Q7af16	escherichia	977	36	43.9	326	2	Q45802	BACVU	Q45802	bacteroides
905	36	43.9	212	2	Q54550	HABIN	Q54550	haemophilus	978	36	43.9	327	2	Q4DVZ0	TRYCR	Q4dvz0	trypanosoma
906	36	43.9	214	2	Q3MPV5	CANAL	Q3mpv5	candida alb	979	36	43.9	327	2	Q4N945	THEPA	Q4n945	theileria p
907	36	43.9	214	2	Q5AFH0	CANAL	Q5afh0	candida alb	980	36	43.9	329	2	Q3JT23	BURP1	Q3jt23	burkholderi

981 36 43.-9 331 2 Q3EXT7_BACTI Q3ext7 bacillus th
 982 36 43.-9 335 1 Q18363_CAEL Q18363 pseudonabdi
 983 36 43.-9 338 1 AMIF_PSSM Q887d9 pseudomonas
 984 36 43.-9 338 2 Q4XA2_PSEU2 Q4xa2 pseudomonas
 985 36 43.-9 342 2 Q4B4K4_9BURK Q4b4k4 polaromonas
 986 36 43.-9 343 2 Q3FVW2_9BURK Q3fvw2 rhodoferrax
 987 36 43.-9 344 2 Q2X9S7_PSEPU Q2x9s7 pseudomonas
 988 36 43.-9 344 2 Q88Q56_PSEPU Q88q56 pseudomonas
 989 36 43.-9 345 2 Q46D88_METBA Q46d88 methanosarc
 990 36 43.-9 345 2 Q8TUM6_METAC Q8tum6 methanosarc
 991 36 43.-9 345 2 Q3F5E9_9BURK Q3f5e9 burkholderi
 992 36 43.-9 345 2 Q40H34_JANNASCHIA Q40h34 jannaschia
 993 36 43.-9 346 2 Q8XYM5_RALSO Q8xym5 ralstonia
 994 36 43.-9 347 2 Q2X7Q0_9GAMM Q2x7q0 shewanella
 995 36 43.-9 347 2 Q3Q5X6_9GAMM Q3q5x6 shewanella
 996 36 43.-9 348 2 Q4XLK2_PLACH Q4xlk2 plasmodium
 997 36 43.-9 348 2 Q3FG54_9BURK Q3fgz4 burkholderi
 998 36 43.-9 348 2 Q44XX9_9BURK Q44xx9 burkholderi
 999 36 43.-9 348 2 Q4BA78_BURVI Q4ba78 burkholderi
 1000 36 43.-9 348 2 Q4LR32_9BURK Q4lr32 burkholderi

ALIGNMENTS

RESULT 1
 HPSE_HUMAN STANDARD; PRT; 543 AA.
 AC Q9Y251; Q53GE5; Q9UL39;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-NOV-1999, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
 DE subunit].
 GN Name:HPSP; Synonyms:HEP, HPA, HPA1, HPRI, HPSE1, HSE1;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
 RA Seddon A.P., Giorgio N.A., Bohlen P.;
 RT "Cloning and functional expression of a human heparanase gene.";
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 [2]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.
 RC TISSUE=Embryonic fibroblast;
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
 RA Toyoshima M., Nakajima M.;
 RT "Human heparanase: Purification, characterization, cloning, and
 RT expression.";
 RL J. Biol. Chem. 274:24153-24160(1999).
 [3]
 RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
 RP PubMed=10395325; DOI=10.1038/10518;
 RX Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
 RA Spector L., Pecker I.;
 RT "Mammalian heparanase: gene cloning, expression and function in tumor
 RT progression and metastasis.";
 RL Nat. Med. 5:793-802(1999).
 [4]
 RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 [5]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage.";
 RL Glycobiology 10:467-475(2000).
 [6]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis.";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
 [7]
 RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Parekh R., Stubberfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpal) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435(2003).
 [8]
 RN NUCLEOTIDE SEQUENCE [MRNA].
 RA Pinhal M.A., Semedo P.;
 RT "Cloned heparanase from MCF-7 cells.";
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 [9]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
 RA Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 [10]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Diatchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Maman A., Rodrigues S., Sanchez A.,
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [11]
 RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RP PubMed=11123890; DOI=10.1021/bi002080p;
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Gready J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic
RL endoglycosidase heparanase.";
RN Biochemistry 39:15659-15667(2000).
[12]
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
RN ASN-217; ASN-238 AND ASN-459.
RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
RA Simizu S., Ishida K., Wierzbicka M.K., Osada H.;
RT "Secretion of heparanase protein is regulated by glycosylation in
RT human tumor cell lines.";
RL J. Biol. Chem. 279:2697-2703(2004).
[13]
RN SUBCELLULAR LOCATION.
RX PubMed=15292202; DOI=10.1074/jbc.M402131200;
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
RT "Heparanase uptake is mediated by cell membrane heparan sulfate
RT proteoglycans.";
RL J. Biol. Chem. 279:44084-44092(2004).
[14]
RN SUBCELLULAR LOCATION, PROCESSING, AND SUBCELLULAR LOCATION.
RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
RT "Heparanase processing by lysosomal/endosomal protein preparation.";
RL FEBS Lett. 579:2334-2338(2005).
[15]
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
RA Aboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
Elgavish S., Peretz T., Vlodavsky I.;
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
RT human proheparanase.";
RL J. Biol. Chem. 280:13568-13575(2005).
[16]
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
RX PubMed=15760902; DOI=10.1074/jbc.M41454200;
RA Levy-Adam F., Aboud-Jarrous G., Guerrini M., Beccati D.,
Vlodavsky I., Ilan N.;
RT "Identification and characterization of heparin/heparan sulfate
RT binding domains of the endoglycosidase heparanase.";
RL J. Biol. Chem. 280:20457-20466(2005).
[17]
RN VARIANT SER-260.
RX PubMed=15334672;
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
Shi L.H., Shen H.W., Xu B.;
RT "Heparanase mRNA expression and point mutation in hepatocellular
RT carcinoma.";
RL World J. Gastroenterol. 10:2795-2799(2004).
CC -1- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
CC lower extent, by heparin and sulfamin and activated by calcium and
CC magnesium (By similarity).
CC -1- BIOPHYSICOCHEMICAL PROPERTIES:
CC pH dependence:
CC Optimum pH is 4-6;
CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products.
CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium.
CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
CC weakly expressed in lymph node, thymus, peripheral blood
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.
CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme.
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.
Query Match 100.0%; Score 82; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SWELGNPNPSFLKKA 15
DB 219 SWELGNPNPSFLKKA 233
[1]
RESULT 2
HPSE_MOUSE
ID HPSE_MOUSE STANDARD; PRT; 535 AA.
AC Q6YG21; Q8K3K3;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=Hps; Synonyms=Hpa;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=JUL/J; TISSUE=Spleen;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RL Nat. Med. 5:803-809(1999).
[2]
RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND
RP SUBUNITS.
RC STRAIN=FVB; TISSUE=Embryo;
RX MEDLINE=2460766; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
RT "Cloning, expression, and purification of mouse heparanase.";
RL Protein Expr. Purif. 26:425-431(2002).
[3]
RN NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,
RA Li J.-P.;
RT "Processing of macromolecular heparin by heparanase.";
RL J. Biol. Chem. 278:35152-35158(2003).
[4]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
RX PubMed=16141072; DOI=10.1126/science.11112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilm M., Wang L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,

OX	NCBI_TaxID=134510;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC	[2]
RN	NUCLEOTIDE SEQUENCE.
RP	TISSUE=Kidney;
RC	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC	-----
CC	Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
DR	EMBL; AM085494; CAJ30021.1; -; mRNA.
SQ	SEQUENCE 558 AA; 62737 MW; 07BAF8F55849EEB7 CRC64;

Query Match 92.7%; Score 76; DB 2; Length 558;	
Best Local Similarity 93.3%; Pred. No. 0.00029;	
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 SWELGNEPNSFLKKA 15
Db	250 SWELGNEPNSFWCKA 264
RESULT 5	
Q333X6 SPAJD	PRELIMINARY; PRT; 574 AA.
ID	Q333X6
AC	Q333X6
DT	06-DSC-2005, integrated into UniProtKB/TrEMBL.
DT	06-DSC-2005, sequence version 1.
DT	07-FEB-2006, entry version 3.
DE	Heparanase.
GN	Name=hpa;
OS	Spalax judeai (Blind subterranean mole rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Palacidae; Spalacinae; Spalax.
OX	NCBI_TaxID=134510;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC	[2]
RN	NUCLEOTIDE SEQUENCE.
RP	TISSUE=Kidney;
RC	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC	-----
CC	Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
DR	EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ	SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match 92.7%; Score 76; DB 2; Length 574;	
Best Local Similarity 93.3%; Pred. No. 0.0003;	
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 SWELGNEPNSFLKKA 15
Db	250 SWELGNEPNSFWCKA 264

```

RESULT 6
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FID19DCBABB99DE CRC64;

Query Match 92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 250 SWELGNEPNSFWKKA 264
|||||

RESULT 7
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 49EBEPEC7D0BCB34 CRC64;

Query Match 92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 250 SWELGNEPNSFWKKA 264
|||||

RESULT 8
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match 92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 250 SWELGNEPNSFWKKA 264
|||||

RESULT 9
HPSE CHICK STANDARD; PRT; 523 AA.
AC Q90VK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462000;
RA Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
```

RT "Expression pattern and secretion of human and chicken heparanase are determined by their signal peptide sequence.";

RL J. Biol. Chem. 276:29178-29187(2001).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans (By similarity).
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
 CC (By similarity).

CC -!- PTM: N-glycosylated (By similarity).
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----
 DR EMBL; AY037007; AAK82648.1; -; mRNA.
 DR Ensembl; ENSGALG0000011203; Gallus gallus.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
 FT SIGNAL 1 18 Potential.
 FT CHAIN 19 523 Heparanase.

FT REGION 137 141 /FTid=PRO_0000042259.
 FT ACT_SITE 250 260 Heparin/HS-binding (By similarity).
 FT ACT_SITE 204 204 Heparin/HS-binding (By similarity).
 FT ACT_SITE 323 323 Nucleophile (Potential).
 FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 523 AA; 58386 MW; 8BE0B7B18C9BF881 CRC64;

Query Match 89.0%; Score 73; DB 1; Length 523;
 Best Local Similarity 86.7%; Pred. No. 0.00092;
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNPNPNSFLKKA 15

|||||

DB 198 SWELGNPNPNSFRKKS 212

RESULT 10

HPSE BOVIN

ID HPSE BOVIN STANDARD; PRT; 545 AA.

AC Q9MY0;

DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.

DT 01-JUN-2001, sequence version 2.

DT 07-MAR-2006, entry version 15.

DE Heparanase precursor (EC 3.2.2.-) [Contains: Heparanase 8 kDa subunit;

DE Heparanase 50 kDa subunit].

GN Name=HPSE;

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;

OC Pecora; Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.

RC TISSUE=Placenta;

RX MEDLINE=21176669; PubMed=11277877;

RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;

RT "Expression of heparanase mRNA in bovine placenta during gestation.";

RL Reproduction 121:573-580(2001).

CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Contributes to metastasis and angiogenesis
 CC (By similarity).

CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
 CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
 CC magnesium (By similarity).

CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunit, the proteolytic products (By similarity).
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium
 CC (By similarity).
 CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
 CC kidney, lung, spleen and uterus.
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
 CC to be essential for its solubility (By similarity).

CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----
 DR EMBL; AF281160; AAF87301.2; -; mRNA.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
 FT SIGNAL 1 37 By similarity.

FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).

FT PROPEP 112 159 Linker peptide.

FT CHAIN 160 545 Heparanase 50 kDa subunit (By similarity).

FT REGION 160 164 /FTid=PRO_0000042256.

FT REGION 272 282 Heparin/HS-binding (Potential).

FT ACT_SITE 227 227 Proton donor (Potential).

FT ACT_SITE 345 345 Nucleophile (Potential).

FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).

SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;

Query Match 89.0%; Score 73; DB 1; Length 545;

Best Local Similarity 86.7%; Pred. No. 0.00097;

Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNPNPNSFLKKA 15

|||||

DB 221 SWELGNPNPNSFORKA 235

RESULT 11

Q4TGC8 TETNG

ID Q4TGC8 TETNG PRELIMINARY; PRT; 255 AA.

AC Q4TGC8;

DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.

DT 19-JUL-2005, sequence version 1.

DT 07-FEB-2006, entry version 4.

DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.

DE (Fragment).

GN ORFNames=GSTENG00001168001;

OS Tetraodon nigroviridis (Green puffer).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

OC Tetraodontidae; Tetraodontidae; Tetraodon.

OX NCBI_TaxID=99883;

RN [1]

RP NUCLEOTIDE SEQUENCE.

```

RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Croliius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAEA01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

Query Match 87.8%; Score 72; DB 2; Length 255;
Best Local Similarity 86.7%; Pred. No. 0.0062;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 54 SWELGNEPNSYEKKA 58
|||||||: |||

RESULT 12
ID Q4SYF6_TETNG PRELIMINARY; PRT; 533 AA.
AC Q4SYF6;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00010356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OC NCBI_TaxID=99883;
OX [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Croliius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
```

```

RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAR01012073; CAF94326.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 533
SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;

Query Match 87.8%; Score 72; DB 2; Length 533;
Best Local Similarity 86.7%; Pred. No. 0.0014;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 174 SWELGNEPNSYEKKA 188
|||||||: |||

RESULT 13
ID HPSE2_HUMAN STANDARD; PRT; 592 AA.
AC Q8WQ02; Q5VUH4; Q5VUH5; Q5VUH6; Q8WQ01; Q9HB37; Q9HB38; Q9HB39;
DT 25-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 25-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 16.
DE Heparanase-2 (EC 3.2.-.-) (Hpa2).
GN Name=HPSE2; Synonyms=HPA2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 3 AND 4), TISSUE SPECIFICITY,
RP AND SUBCELLULAR LOCATION.
RC TISSUE=Heart;
RX MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbrc.2000.3586;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and expression profiling of Hpa2, a novel mammalian
RT heparanase family member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TYR-579.
RX PubMed=15164054; DOI=10.1038/nature02462;
RA Deloukas P., Earthrowl M.E., Grafham D.V., Rubinfeld M., French L.,
RA Steward C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,
RA Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.L.,
RA Taylor A., Battles J., Bird C.P., Ainscough R., Almeida J.P.,
RA Ashwell R.I.S., Ambrose K.D., Babbage A.K., Baguley C.L., Bailey J.,
RA Brown J.E., Bates K., Beasley H., Bray-Allen S., Brown A.J.,
RA Brown J.Y., Burford D.C., Burrill W., Burton J., Cahill P., Camire D.,
RA Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Corby N., Coulson A., Dhani P., Dutta I., Dunn M., Faulkner L.,
RA Frankish A., Frankland J.A., Garner P., Garnett J., Gribble S.,
RA Griffiths C., Grocock R., Gustafson E., Hammond S., Harley J.L.,
RA Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.,
RA Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,
RA Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,
```



```

Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 266

RESULT 14
Q2MIH9 HUMAN
ID Q2MIH9 HUMAN PRELIMINARY; PRT; 592 AA.
AC Q2MIH9
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 1.
DE Heparanase 2.
GN Name=HPSE2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RG NIH MGC Project;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL: BC112356; AAI12357.1; -; mRNA.
SQ SEQUENCE 592 AA; 66610 MW; 94689B1C2A74359F CRC64;

Query Match 70.7%; Score 58; DB 2; Length 592;
Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 266

RESULT 15
Q4TB80 TETNG
ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.

```

```

DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG00003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionet C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype."
RL Nature 431:945-957 (2004).
[2]
RN NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL: CAAB01007180; CAP98952.1; -; Genomic DNA.
FT NON TER. 597 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match 70.7%; Score 58; DB 2; Length 597;
Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 268

RESULT 16
Q8TI08 BOMMO
ID Q8TI08 BOMMO PRELIMINARY; PRT; 515 AA.
AC Q8TI08;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Heparanase-like protein.
GN Name=Bmhepa;
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=p50; TISSUE=Posterior silk gland;
RA Koike Y., Mita K., Suzuki M.G., Maeda S., Abe H., Osoegawa K.,
RA deJong P.J., Shimada T.;
RT "Genomic sequence of a 320-kb segment of the Z chromosome of Bombyx
RT mori containing a kettin ortholog."
RL Mol. Genet. Genomics 269:137-149 (2003).

```

```

CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB079860; BAB85191.1; -; Genomic DNA.
DR EMBL; AB090307; BAC10612.1; -; Genomic DNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 515 AA; 59770 MW; FB8100AB6EDDADB CRC64;

Query Match      69.5%; Score 57; DB 2; Length 515;
Best Local Similarity 90.0%; Pred. No. 0.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEPNSF 11
|:|||||
Db 194 WQLGNEPNSF 203

RESULT 17
BGLR_CANFA
ID BGLR_CANFA STANDARD; PRT; 651 AA.
AC O18835;
DT 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.
DT 01-JAN-1998, sequence version 1.
DT 21-FEB-2006, entry version 40.
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=GUSB;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND VARIANT MPS VII HLS-166.
RA MEDLINE=38190525; PubMed=9521879; DOI=10.1006/geno.1997.5189;
RA Ray J., Bouvet A., Desanto C., Fyfe J.C., Xu D., Wolfe J.H.,
RA Aguirre G.D., Patterson D.F., Haakins M.E., Henthorn P.S.;
RT "Cloning of the canine beta-glucuronidase cDNA, mutation
RT identification in canine MPS VII, and retroviral vector-mediated
RT correction of MPS VII cells.";
RL Genomics 48:248-253(1998).
CC -!- FUNCTION: Plays an important role in the degradation of dermatan
CC and keratan sulfates [By similarity].
CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
CC + an alcohol.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosome.
CC -!- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis
CC type VII (MPS VII), an inherited disease reported in humans, mice,
CC cats, and dogs.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF019759; AAC48809.1; -; mRNA.
DR HSP; P08236; 1BHG.
DR SMR; O18835; 1BHG.
DR Ensembl; ENSCAFG00000010193; Canis familiaris.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2IG.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR InterPro; IPR006104; Glyco_hydro_S_bd.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
DR Signal.
FT SIGNAL 1 22 By similarity.

```

```

FT CHAIN 23 651 Beta-glucuronidase.
FT ACT_SITE 450 450 Proton donor (By similarity).
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).
FT VARIANT 166 166 R -> H (in MPS VII; loss of activity).
SQ SEQUENCE 651 AA; 74433 MW; E8991B1E65C60120 CRC64;

Query Match      63.4%; Score 52; DB 1; Length 651;
Best Local Similarity 64.3%; Pred. No. 5.9;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
|:|||||
Db 445 WSVANEPTSFLKPA 458

RESULT 18
Q4FAT7_PIG
ID Q4FAT7_PIG PRELIMINARY; PRT; 652 AA.
AC Q4FAT7;
DT 30-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 30-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Beta-glucuronidase.
GN Name=GUSB;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Beck J., Knorr C., Brenig B.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
CC + an alcohol.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ095863; AA203639.1; -; Genomic DNA.
DR SMR; Q4FAT7; 24-632.
DR GO; GO:0005764; C:lysosome; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl...; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2IG.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR Lysosome.
SQ SEQUENCE 652 AA; 74710 MW; 7B9386564DF2CA7 CRC64;

Query Match      62.2%; Score 51; DB 2; Length 652;
Best Local Similarity 57.1%; Pred. No. 8.9;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
|:|||||
Db 446 WSVANEPTSFLKPA 459

RESULT 19
BGLR_FELCA
ID BGLR_FELCA STANDARD; PRT; 651 AA.
AC O97524;
DT 15-JUL-1999, integrated into UniProtKB/Swiss-Prot.

```

DT 01-MAY-1999, sequence version 1.
DT 21-FEB-2006, entry version 39.
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=GUSB;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
OC Felinae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND VARIANT MPS VII LYS-351.
RC TISSUE=Liver;
RX MEDLINE=99296826; PubMed=10366443; DOI=10.1006/geno.1999.5825;
RA Fyfe J.C., Kurzhals R.L., Lassaune M.E., Henthorn P.S., Alur P.R.,
RA Wang P., Wolfe J.H., Giger U., Haskins M.E., Patterson D.F., Sun H.,
RA Jain S., Yuhki N.;
RT "Molecular basis of feline beta-glucuronidase deficiency: an animal
RT model of mucopolysaccharidosis VII.";
RL Genomics 58:121-128(1999).
CC -1- FUNCTION: Plays an important role in the degradation of dermatan
CC and keratan sulfates.
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
CC + an alcohol.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Lysosome.
CC -1- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis
CC type VII (MPS VII), an inherited disease reported in humans, mice,
CC cats, and dogs.
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AF012423; AAD01498.1; -; mRNA.
DR EMBL; AF012424; AAD01499.1; -; mRNA.
DR HSSP; P08236; 1BHG.
DR SNR; O97524; 22-631.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2IG.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR InterPro; IPR006104; Glyco_hydro_2_bd.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
KW Signal.
FT SIGNAL 1 22 By similarity.
FT CHAIN 23 651 Beta-glucuronidase.
FT /FTID=PRO_0000012160.
FT ACT_SITE 450 450 Proton donor (By similarity).
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 530 530 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 351 351 E -> K (in MPS VII).
FT VARIANT 651 AA; 74610 MW; 2AE30884B70D4232 CRC64;
SQ SEQUENCE 651 AA; 74610 MW; 2AE30884B70D4232 CRC64;
Query Match 61.0%; Score 50; DB 1; Length 651;
Best Local Similarity 64.3%; Pred. No. 13;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKKA 15
| : ||| |||||
DB 445 WSVANEPASFLKPA 458

RESULT 20
Q6FYV6 BARQU
ID Q6FYV6 BARQU PRELIMINARY; PRT; 581 AA.
AC Q6FYV6;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
GN OrderedLocusNames=BQ10670;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Toulouse;
RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
RA Almark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Canbaeck B., Eriksson A.-S., Naesslund A.K., Handley S.A., Huvet M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; BX897700; CAF26534.1; -; Genomic DNA.
SQ SEQUENCE 581 AA; 65072 MW; 146DEC3FF6339D24 CRC64;
Query Match 59.8%; Score 49; DB 2; Length 581;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 SWELGNEPNSFLKKA 15
| : ||| |||||
DB 307 SWEVANHPTFSKLA 321

RESULT 21
Q2TYD4 ASPOR
ID Q2TYD4 ASPOR PRELIMINARY; PRT; 260 AA.
AC Q2TYD4;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 24-MAR-2006, entry version 3.
DE Predicted protein.
GN ORFNames=AO090103000292;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RX PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA Kuhara S., Ogasawara N., Kikuchi H.;
RT "Genome sequencing and analysis of Aspergillus oryzae.";
RL Nature 438:1157-1161(2005).
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AP007174; BAE65739.1; -; Genomic DNA.
SQ SEQUENCE 260 AA; 28377 MW; FA60BE90767F5982 CRC64;

```
Query Match          58.5%; Score 48; DB 2; Length 260;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
   :||:||||:
Db 168 AWEIGNEPDNY 178

RESULT 22
ID Q64Y35_BACFR PRELIMINARY; PRT; 435 AA.
AC Q64Y35;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Endo-1,4-beta-mannosidase.
GN OrderedLocusNames=BF0840;
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
[1]
RN SMART; SM00490; HELIC; 1.
RC PUBMED=15466707; DOI=10.1073/pnas.0404172101;
RA Kuwahara T., Yamashita A., Hirakawa H., Nakayama H., Toh H., Okada N.,
RA Kuhara S., Hattori M., Hayashi T., Ohnishi Y.;
RT "Genomic analysis of Bacteroides fragilis reveals extensive DNA
RT inversions regulating cell surface adaptation.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14919-14924(2004).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL; AP006841; BAD47591.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 435 AA; 48916 MW; 34D8423F8DDE86CB CRC64;

Query Match          58.5%; Score 48; DB 2; Length 435;
Best Local Similarity 63.6%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
   :||:||||:
Db 215 SWQIGNEPRPF 225

RESULT 23
ID Q3IKG9_PSEHT PRELIMINARY; PRT; 1297 AA.
AC Q3IKG9;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Helicase, ATP-dependent (EC 3.6.1.15).
GN Name=hrpA; OrderedLocusNames=PSHA1144; ORFNames=PSHA1144;
OS Pseudalteromonas haloplanktis (strain TAC 125).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudalteromonadaceae; Pseudalteromonas.
OX NCBI_TaxID=326442;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP PubMed=16169927; DOI=10.1101/gr.4126905;
RA Medigue C., Krin E., Pascal G., Barbe V., Bernsel A., Bertin P.N.,
RA Cheung F., Cruveiller S., D'Amico S., Duilio A., Fang G., Feller G.,
RA Ho C., Mangelot S., Marino G., Nilsson J., Parrilli E., Rocha E.P.C.,
RA Rouy Z., Sekowska A., Tutino M.L., Vallenet D., von Heijne G.,
RA Danchin A.;
RT "Coping with cold: the genome of the versatile marine Antarctic
RT bacterium Pseudalteromonas haloplanktis TAC125.";
RL Genome Res. 15:1325-1335(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC

Query Match          57.3%; Score 47; DB 2; Length 1297;
Best Local Similarity 53.8%; Pred. No. 98;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKK 14
   :|:||||:
Db 1019 WDFGELPNSYVKK 1031

RESULT 24
ID Q3CKA3_THEET PRELIMINARY; PRT; 627 AA.
AC Q3CKA3;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Glycoside hydrolase, family 5.
GN ORFNames=Teth39DRAFT_0967;
OS Thermoanaerobacter ethanolicus ATCC 33223.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=340099;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=ATCC 33223;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Thermoanaerobacter
RT ethanolicus 39E.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33223;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Thermoanaerobacter ethanolicus
RT 39E.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
```

```
CC -----
DR EMBL; AAKQ01000001; EAO65730.1; -; Genomic_DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
KW Hydrolase.
SQ SEQUENCE 627 AA; 72436 MW; BD4FB389872AB8FC CRC64;

Query Match 56.1%; Score 46; DB 2; Length 627;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
Db 139 WDLSPNDNVVK 150
|:|:|:|:|:|:|

RESULT 25
Q8RDH1 THETN PRELIMINARY; PRT; 637 AA.
AC Q8RDH1;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-FEB-2006, entry version 18.
DE Hypothetical protein.
GN OrderedLocusNames=TT00061;
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336; DOI=10.1101/gr.219302;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of the T. tengcongensis genome.";
RL Genome Res. 12:1689-700(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AEO12980; AAM23368.1; -; Genomic_DNA.
DR BIOCYC; TTEN119072:TT00061-MONOMER; -;
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; Cellulase; 1.
KW Complete proteome.
SQ SEQUENCE 637 AA; 74206 MW; 619E35812AC83EBB CRC64;

Query Match 56.1%; Score 46; DB 2; Length 637;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
Db 149 WDLSPNDNVVK 160
|:|:|:|:|:|:|

RESULT 26
Q3E4N3 CHLAU PRELIMINARY; PRT; 651 AA.
AC Q3E4N3;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Glycoside hydrolase, family 5.
GN ORFNames=CaurDRAFT_0879;
OS Chloroflexus aurantiacus J-10-fl.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.
OX NCBI_TaxID=324602;
```

```
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=J-10-fl;
RC US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Chloroflexus
RT aurantiacus J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=J-10-fl;
RC US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Chloroflexus aurantiacus
RT J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[3]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=J-10-fl;
RC US DOE Joint Genome Institute;
RA DOE Joint Genome Institute;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
[4]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=J-10-fl;
RC US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AARH02000003; EAO60769.1; -; Genomic_DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
KW Hydrolase.
SQ SEQUENCE 651 AA; 71446 MW; 2737C9294AC882FE CRC64;

Query Match 56.1%; Score 46; DB 2; Length 651;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEPNS 10
Db 350 WELGNEPDA 358
|:|:|:|:|:|:|

RESULT 27
Q2U2I3 ASPOR PRELIMINARY; PRT; 373 AA.
ID Q2U2I3 ASPOR PRELIMINARY; PRT; 373 AA.
AC Q2U2I3;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-MAR-2006, entry version 3.
DE Endo-beta-mannanase.
GN ORFNames=AO090038000444;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RX PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
```

RA Bhatnagar A., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA Hosoyana D., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato I., Kin I., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi T., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA Kuhara S., Ogasawara N., Kikuchi H.,
RA "Genome sequencing and analysis of *Aspergillus oryzae*.";
RT Nature 438:1157-1161(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007169; BAB64232.1; -; Genomic DNA.
SQ SEQUENCE 373 AA; 41126 MW; 3AD40B2A6AFC7486 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 373;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNP 8
Db 190 AWELGNP 197
:|||||

RESULT 28
Q5B833 EMENI PRELIMINARY; PRT; 409 AA.
AC Q5B833
DT 26-APR-2005, integrated into UniProtKB/TrEMBL.
DT 26-APR-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Hypothetical protein.
GN ORFNames=AN3297.2;
OS *Aspergillus nidulans* FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC PubMed=16372000; DOI=10.1038/nature04341;
RX Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R.,
RA Batzoglou S., Lee S.-I., Basturkmen M., Spevak C.C., Clutterbuck J.,
RA Kapitonov V., Jurka J., Scanzocchio C., Farman M., Butler J.,
RA Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C.,
RA Boucher C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S.,
RA Doonan J.H., Yu J., Vienken K., Pain A., Freitag M., Selker E.U.,
RA Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T.,
RA Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W.,
RA Caddick M., Hynes M., Paoletti M., Fischer R., Miller B.L., Dyer P.S.,
RA Sachs M.S., Osmari S.A., Birren B.W.;
RA "Sequencing of *Aspergillus nidulans* and comparative analysis with A.
RT fumigatus and A. *oryzae*.";
RL Nature 438:1105-1115(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAC001000035; EAA63285.1; -; Genomic DNA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; Cellulase; 1.
KW Hypothetical protein.
SQ SEQUENCE 409 AA; 46168 MW; 93AB6EC788222720 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 409;
Best Local Similarity 87.5%; Pred. No. 61;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNP 8
Db 220 AWELGNP 227
:|||||

RESULT 29
Q99036 TRIRE PRELIMINARY; PRT; 437 AA.
AC Q99036
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 26.
DE Beta-mannase precursor.
OS *Trichoderma reesei* (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RutC30;
RA Stalbrand H., Saloheimo A., Vehmaanpera J., Penttila M.;
RT "cDNA encoding *Trichoderma reesei* beta-mannanase.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; L25310; AAA34208.1; -; mRNA.
DR PDB; 1QNP; X-ray; A=28-371.
DR PDB; 1QNP; X-ray; A=28-371.
DR PDB; 1QNP; X-ray; A=28-371.
DR PDB; 1QNP; X-ray; A=28-371.
DR PDB; 1QNR; X-ray; A=28-371.
DR PDB; 1QNS; X-ray; A=28-371.
DR GO; GO:0005576; C:extracellular region; IEA.
DR GO; GO:0030248; F:cellulose binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000254; CBD fun.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00150; Cellulase; 1.
DR ProDom; PD001821; CBD fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 28 437 beta-mannase.
SQ SEQUENCE 437 AA; 47053 MW; 17513DADE12654A7 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 437;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNP 8
Db 190 AWELGNP 197
:|||||

RESULT 30
Q3DZN2 CHLAU PRELIMINARY; PRT; 490 AA.
AC Q3DZN2
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=CaurDRAFT_3045;
OS *Chloroflexus aurantiacus* J-10-fl.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.
OX NCBI_TaxID=324602;
RN [1]

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Chloroflexus
RT aurantiacus J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Chloroflexus aurantiacus
RT J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
US DOE Joint Genome Institute;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
[4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AA02000021; EA058861.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 490 AA; 53840 MW; 103222A5536FE211 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 490;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEP 9
Db 163 WELGNEPD 170

RESULT 31
Q5CW71 CRYPV PRELIMINARY; PRT; 538 AA.
AC Q5CW71;
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.
DT 12-APR-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE DNA primase large subunit (Fragment).
GN ORFNames:cgd8_1410;
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Iowa type II;
RX PubMed=15044751; DOI=10.1126/science.1094786;
RA Abrahamson M.S., Templeton T.J., Enomoto S., Abrahante J.E., Zhu G.,
RA Lantto C.A., Deng M., Liu C., Widmer G., Tzipori S., Buck G.A., Xu P.,
RA Bankier A.T., Dear P.H., Konfortov B.A., Spriggs H.F., Iyer L.,
RA Anantharaman V., Aravind L., Kapur V.;
RT "Complete genome sequence of the apicomplexan, Cryptosporidium
RT parvum.";
RL Science 304:441-445(2004).
CC -!- CAUTION: The sequence shown here is derived from an

```

```

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAE0100003; EAK9678.1; -; Genomic_DNA.
DR GO; GO:0005658; C:alpha DNA polymerase:primase complex; IEA.
DR GO; GO:0003896; F:DNA primase activity; IEA.
DR GO; GO:0006269; P:DNA replication, synthesis of RNA primer; IEA.
DR InterPro; IPR007238; DNA_primase_lrg.
DR Pfam; PF04104; DNA_primase_lrg; I.
FT NON_TER 1
SQ SEQUENCE 538 AA; 63237 MW; 9AE256D7E36927C6 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 538;
Best Local Similarity 58.3%; Pred. No. 82;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 LGNEPNSFLKKA 15
Db 514 IGHNFNSYFKS 525

RESULT 32
BGLR MOUSE
ID BGLR MOUSE STANDARD; PRT; 648 AA.
AC P12265; Q61601; Q64473; Q64474;
DT 01-OCT-1989, integrated into UniProtKB/Swiss-Prot.
DT 01-OCT-1989, sequence version 1.
DT 21-FEB-2006, entry version 59.
DE Beta-glucuronidase precursor [EC 3.2.1.31].
GN Name=Gusb; Synonyms=Gus, Gus-s;
OS Mus musculus [Mouse].
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88085188; PubMed=2891607;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Elliott R.W., Pazik J.,
RA Hohman C., Korfhagen T.R., Ganschow R.E.;
RT "DNA sequence variation within the beta-glucuronidase gene complex
RT among inbred strains of mice.";
RL Genomics 1:145-152(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88284700; PubMed=3397060;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Ganschow R.E.;
RA Gallagher P.M., D'Amore M.A., Korfhagen T.R., Ganschow R.E.;
RT "The complete nucleotide sequence of murine beta-glucuronidase mRNA
RT and its deduced polypeptide.";
RL Genomics 2:215-219(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89062453; PubMed=3196706;
RA D'Amore M.A., Gallagher P.M., Korfhagen T.R., Ganschow R.E.;
RT "Complete sequence and organization of the murine beta-glucuronidase
RT gene.";
RL Biochemistry 27:7131-7140(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C3H/HeJ and YBR; TISSUE=Sperm;
RX MEDLINE=89384641; PubMed=2779578;
RA Wawrzyniak C.J., Gallagher P.M., D'Amore M.A., Carter J.E., Lund S.D.,
RA Rinchik E.M., Ganschow R.E.;
RT "DNA determinants of structural and regulatory variation within the
RT murine beta-glucuronidase gene complex.";
RL Mol. Cell. Biol. 9:4074-4078(1989).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88216590; PubMed=2835664;
RA Funkenstein B., Leary S.L., Stein J.C., Catterall J.F.;

```


RT "Genomic organization and sequence of the Gus-s alpha allele of the
 RL murine beta-glucuronidase gene.";
 RN Mol. Cell. Biol. 8:1160-1168(1988).
 RP NUCLEOTIDE SEQUENCE OF 593-648, VARIANT ARG-642, AND SUBCELLULAR
 RP LOCATION.
 RX MEDLINE=90368633; PubMed=2394691;
 RA Li H., Takeuchi K.H., Manly K., Chapman V., Swank R.T.;
 RT "The propetide of beta-glucuronidase. Further evidence of its
 RT involvement in compartmentalization of beta-glucuronidase and sequence
 RT similarity with portions of the reactive site region of the serpin
 RT superfamily.";
 RL J. Biol. Chem. 265:14732-14735(1990).
 CC -!- FUNCTION: Plays an important role in the degradation of dermatan
 CC and keratan sulfates.
 CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
 CC + an alcohol.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- SUBCELLULAR LOCATION: Lysosomal. A small proportion is found in
 CC the endoplasmic reticulum.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC -----
 DR EMBL: J03047; AAA37696.1; -; mRNA.
 DR EMBL: J02836; AAA98623.1; -; Genomic_DNA.
 DR EMBL: M63836; AAA63309.1; -; mRNA.
 DR EMBL: M28540; AAA63307.1; -; mRNA.
 DR EMBL: M28541; AAA63308.1; -; mRNA.
 DR EMBL: M19279; AAA37697.1; -; mRNA.
 DR PIR: A32576; A32576.
 DR HSSP: P08236; 1BHG.
 DR SMR: P12265; 22-628.
 DR Ensembl: ENSMUSG00000025534; Mus musculus.
 DR MGI: MGI:95872; Gusb.
 DR InterPro: IPR006101; Glyco_hydro_2.
 DR InterPro: IPR006102; Glyco_hydro_2lg.
 DR InterPro: IPR006103; Glyco_hydro_2TIM.
 DR InterPro: IPR006104; Glyco_hydro_s_bd.
 DR Pfam: PF00703; Glyco_hydro_2; 1.
 DR Pfam: PF02836; Glyco_hydro_2_C; 1.
 DR Pfam: PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS: PR00132; GLHYRIASE2.
 DR PROSITE: PS00719; GLYCOSYL HYDROL_F2_1; 1.
 DR PROSITE: PS00608; GLYCOSYL HYDROL_F2_2; 1.
 DR KEGG: 05110; Glycosyl hydrolase 2.
 KW Endoplasmic reticulum; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
 KW Polymorphism; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 648
 FT FTID=PRO_0000012162.
 FT Beta-glucuronidase.
 FT Proton donor (By similarity).
 FT N-linked (GlcNAc...) (Potential).
 FT N-linked (GlcNAc...) (Potential).
 FT N-linked (GlcNAc...) (Potential).
 FT N-linked (GlcNAc...) (Potential).
 FT I -> T (in strain: C3H/HeJ).
 FT I -> T (in allele GUS-SA).
 FT D -> G (in strain: YBR and C3H/HeJ).
 FT V -> I (in strain: YBR and C3H/HeJ).
 FT E -> K (in allele GUS-SA).
 FT F -> L (in allele GUS-SA).
 FT G -> R (in allele W26; reduced retention
 FT in the endoplasmic reticulum).
 FT SEQUENCE 648 AA; 74239 MW; 308C65A50B3B96D6 CRC64;
 Query Match 54.9%; Score 45; DB 1; Length 648;
 Best Local Similarity 57.1%; Pred. NO. 1e+02;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 WELGNPNSFLKKA 15
 | : ||| : ||
 Db 442 WSVANEPSSALKPA 455

RESULT 33
 Q3TW82_MOUSE
 ID Q3TW82_MOUSE PRELIMINARY; PRT; 648 AA.
 AC Q3TW82;
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 5.
 DT 07-FEB-2006, entry version 5.
 DE Osteoclast-like cell cDNA, RIKEN full-length enriched library,
 DE clone:I420031k20 product:beta-glucuronidase, full insert sequence
 DE (Osteoclast-like cell cDNA, RIKEN full-length enriched library,
 DE clone:I420006F06 product:beta-glucuronidase, full insert sequence).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Methods Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Ainslie V., Allen J.E.,
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bono H., Chalk A.M.,
 RA Bansal M.P., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Humenicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
 RA Tamajoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome.";
 RL Science 305:1559-1563(2005).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG RIKEN Genome Exploration Research Group, and Genome Science Group

RG (Genome Network Core Team) and the PANTOM Consortium;
 RT "Antisense Transcription in the Mammalian Transcriptome.";
 RN Science 309:1564-1566 (2005).
 RP [4]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nigaki I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii I., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RN Nature 420:563-573 (2002).
 RP [5]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikolaou I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690 (2001).
 RP [6]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630 (2000).
 RP [7]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771 (2000).
 RP [8]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
 CC + an alcohol.

CC -!- SUBUNIT: Homotrimer (By similarity).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NonCommercial License
 CC -----
 CC EMBL; AK159804; BAE35384.1; -; mRNA.
 CC EMBL; AK159163; BAE34865.1; -; mRNA.
 CC GO; GO:0005764; C:lysosome; RCA.
 CC GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; RCA.
 CC GO; GO:0005975; P:carbohydrate metabolism; RCA.
 CC InterPro; IPR006101; Glyco_hydro_2.
 CC InterPro; IPR006102; Glyco_hydro_21g.
 CC InterPro; IPR006103; Glyco_hydro_2TIM.
 CC InterPro; IPR006104; Glyco_hydro_5_bd.
 CC Pfam; PF00703; Glyco_hydro_2; 1.
 CC Pfam; PF02837; Glyco_hydro_2_C; 1.
 CC Pfam; PF02836; Glyco_hydro_2_N; 1.
 CC PRINTS; PR00132; GLHYDRLASE2.
 CC PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 CC PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 CC Lysosome.

SK SEQUENCE 648 AA; 74271 MW; 93F59C49BB186BB5 CRC64;

Query Match Similarity 54.9%; Score 45; DB 2; Length 648;

Best Local Similarity 57.1%; Pred. No. le+02; Mismatches 4; Indels 0; Gaps 0;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15

Db 442 WSVANEPSSALKKA 455

AC Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 07-FEB-2006, entry version 19.

DE Glucuronidase, beta (Bone marrow macrophage cDNA, RIKEN full-length

DE insert sequence) (Osteoclast-like cell cDNA, RIKEN full-length

DE insert sequence) (12 days embryo embryonic body between diaphragm

DE region and neck cDNA, RIKEN full-length enriched library,

DE clone:9430030C21 product:beta-glucuronidase, full insert sequence)

DE (Adult male cecum cDNA, RIKEN full-length enriched library,

DE clone:9130020116 product:beta-glucuronidase, full insert sequence).

GN Name=Gusb;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidea; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bobak S., McEwen P.J., McKernan K.J., Carninci P., Prange C.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smal M.A.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
 RC Director MGC Project;
 RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 RL [3]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
 RC diaphragm region and neck;
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Methods Enzymol. 303:19-44 (1999).
 RN [4]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
 RC diaphragm region and neck;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.L., Dally-Murray B.P., de Bono B., Della Gatta G.,
 RA di Bernardo D., Down T., Engstrom P., Fegiolini M., Faulkner G.,
 RA Fletcher-Hemming P., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollas L., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 RA Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J.,
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 RA Motaghi-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
 RA Petrovsky N., Piazza S., Reid J.F., Ring B.Z., Ringwald M.,
 RA Roost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugiyara K., Sultana R., Takenaka Y., Taki K.,
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,

Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kishimoto T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RT "The transcriptional landscape of the mammalian genome.";
 RL Science 309:1559-1563 (2005).
 RN [5]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
 RC diaphragm region and neck;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG RIKEN Genome Exploration Research Group, and Genome Science Group
 (Genome Network Core Team) and the FANTOM Consortium;
 RT "Antisense Transcription in the Mammalian Transcriptome.";
 RL Science 309:1564-1566 (2005).
 RN [6]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
 RC diaphragm region and neck;
 RX MEDLINE=22354663; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer J.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyszewski-Boris A., Yanagisawa M., Yang L., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 RN [7]

NUCLEOTIDE SEQUENCE.
 RP STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
 RC diaphragm region and neck;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bargh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
RC diaphragm region and neck;
RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu N., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[9]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between
RC diaphragm region and neck;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[10]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, and Cecum;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
Query Match 54.9%; Score 45; DB 2; Length 648;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 WELGNPNSFLKKA 15
| : ||||| :
DB 442 WSNVNEPSSALKPA 455

RESULT 35
Q4IR74 GIBZE PRELIMINARY; PRT; 1687 AA.
AC Q4IR74;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=FG00284.1;
OS Gibberella zeae (Fusarium graminearum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5518;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PH-1 / NRRL 31084;
RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choel Y., Collymore A., Cook A., Cooke P., Corum B., DeArelano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,
RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,
RA Hafez N., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,
RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,
RA Karatas A., Kellis C., Landers T., Levine R., Lindblad-Toh K., Liu G.,

RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,
RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,
RA Menus L., Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C.,
RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,
RA O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,
RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,
RA Rogov P., Roman J., Schauer S., Schupack R., Seaman S., Severy P.,
RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,
RA Stubbs M., Talmas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.S.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AACM01000010; EAA69694.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1687 AA; 189392 MW; 9DFAI3423199466C CRC64;
Query Match 54.9%; Score 45; DB 2; Length 1687;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 SWELGNPNSFLKK 14
| : ||||| :
DB 1106 SWILGNBPXPYNE 1119

RESULT 36
Q3D554 STRAG PRELIMINARY; PRT; 90 AA.
AC Q3D554;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=SAN 2117;
OS Streptococcus agalactiae COH1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342616;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=COH1;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dmitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RT "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAJ01000246; EAO74619.1; -; Genomic_DNA.
KW Hypothetical protein.

```
SQ SEQUENCE 90 AA; 10511 MW; 5E2CDE355801D9B4 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLK 14
||| ||| |||
Db 31 SWEKAGELNSLLK 44

RESULT 37
Q3DNV8_STRAG PRELIMINARY; PRT; 90 AA.
AC Q3DNV8;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=SAL_2048;
OS Streptococcus agalactiae 515.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342614;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=515;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Daviden J.T.M., Mora M., Scarselli M., Margarit Y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RT "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AALP01000002; EAO72141.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 90 AA; 10568 MW; 57254EAC5198D9BB CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLK 14
||| ||| |||
Db 31 SWEKAGELNSLLK 44

RESULT 38
Q48S82_STRPM PRELIMINARY; PRT; 90 AA.
AC Q48S82;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Hypothetical protein.
GN OrderedLocNames=M28_Spy1318;
OS Streptococcus pyogenes serotype M28.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
```

```
OX NCBI_TaxID=319700;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MQA6180 / Serotype M28;
RX PubMed=16088825; DOI=10.1086/430618;
RA Green N.M., Zhang S., Porcella S.F., Nagiec M.J., Barbian K.D.,
RA Beres S.B., Lefebvre R.B., Musser J.M.;
RT "Genome sequence of a serotype M28 strain of group A Streptococcus:
RT potential new insights into puerperal sepsis and bacterial disease
RT specificity.";
RL J. Infect. Dis. 192:760-770(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; CP000056; AAX72428.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 90 AA; 10568 MW; 5694EC2EC18B5A2B CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLK 14
||| ||| |||
Db 31 SWEKAGELNSLLK 44

RESULT 39
Q8DX45_STRAS PRELIMINARY; PRT; 90 AA.
AC Q8DX45;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Hypothetical protein SAG2012.
GN OrderedLocNames=SAG2012;
OS Streptococcus agalactiae serotype V.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2603 V/R / ATCC BAA-511 / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AB014282; AAN00871.1; -; Genomic_DNA.
DR TIGR; SAG2012; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 90 AA; 10511 MW; 5E2CDE355801D9B4 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLK 14
||| ||| |||
Db 31 SWEKAGELNSLLK 44
```


CC -|- SIMILARITY: Belongs to the nucleosome assembly protein (NAP)
 CC family.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AF042180; AACG2303.1; -; mRNA.
 CC EMBL; AK053812; BAC35536.1; -; mRNA.
 CC EMBL; BC011213; AAH11213.1; -; mRNA.
 CC Ensembl; ENSMUSG0000047544; Mus musculus.
 CC MGI; MGI:1298395; Teyyl1.
 CC InterPro; IPR002164; NAP family.
 CC PANTHER; PTHR11875; NAP_family; 1.
 CC Pfam; PF00956; NAP; 1.
 KW Nuclear protein.
 FT CHAIN 1 379 Testis-specific Y-encoded-like protein 1.
 FT /FTId=PRO_0000185671.
 SQ SEQUENCE 379 AA; 42994 MW; 92869CAA557FC2D CRC64;
 Query Match 53.7%; Score 44; DB 1; Length 379;
 Best Local Similarity 46.2%; Pred. No. 84;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WELGNPNFLKK 14
 DB 299 WRRGEPQPIRR 311
 RESULT 42
 Q3TKWO_MOUSE PRELIMINARY; PRT; 379 AA.
 AC Q3TKWO;
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
 DT 07-FEB-2006, entry version 1.
 DE Blastocyst blasocyst cDNA, RIKEN full-length enriched library.
 DE clone:11C0014007 product:testis-specific protein, Y-encoded-like, full
 DE insert sequence (Blastocyst blasocyst cDNA, RIKEN full-length
 DE enriched library, clone:11C0020N16 product:testis-specific protein, Y-
 DE encoded-like, full insert sequence) (Osteoclast-like cell cDNA, RIKEN
 DE full-length enriched library, clone:1420019020 product:testis-specific
 DE protein, Y-encoded-like, full insert sequence).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RT Carninci P., Hayashizaki Y.;
 RL "High-efficiency full-length cDNA cloning.";
 RL Methods Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 RA Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 RA di Bernardo D., Down T., Engstrom S., Fagioli M., Faulkner G.,
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,

RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 RA Liumi S., McWilliam S., Madan Babu M., Madera M., Marchionni L., K.,
 RA Matsuda H., Matsuwaga S., Miki H., Mignone F., Miyake S., Morris L.,
 RA Natsugai-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
 RA Petkovsky N., Piazza S., Reid J.F., Ring B.Z., Ringwald M.,
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Seesla L.,
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 RA Sperling S., Stupka E., Sugitara K., Sultana R., Takenaka Y., Taki K.,
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 RA Grimmond S.M., Tesdale R.D., Liu E.T., Brusic V., Quackenbush J.,
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Nimmiya N.,
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
 RA Tagami M., Waki K., Wataniki A., Okamura-Oho Y., Suzuki H., Kawai J.,
 RA Hayashizaki Y.;
 RL "The transcriptional landscape of the mammalian genome.";
 RL Science 309:1559-1563(2005).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX PubMed=16141073; DOI=10.1126/science.1112009;
 RG RIKEN Genome Exploration Research Group, and Genome Science Group
 RG (Genome Network Core Team) and the FANTOM Consortium;
 RT "Antisense Transcription in the Mammalian Transcriptome.";
 RL Science 309:1564-1566(2005).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354583; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Oatso N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kasasawa Y., Kedziński R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reid J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Tesdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,


```

RESULT 44
Q5TFE6 HUMAN PRELIMINARY; PRT; 437 AA.
ID Q5TFE6
AC Q5TFE6;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE OTHUMP00000017052.
GN Name=TSPLY; ORFNames=RP3-486I3.3-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Barlow K.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBSJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AL050331; CAB55883.1; -; Genomic DNA.
DR Ensembl; ENSG00000189241; Homo sapiens.
DR LinkHub; Q5TFE6; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP family.
DR PANTHER; PTHR11875; NAP_Family; 1.
DR Pfam; PF00956; NAP; 1.
SQ SEQUENCE 437 AA; 49192 MW; D651AGFAA7CF811F CRC64;

Query Match 53.7%; Score 44; DB 2; Length 437;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WELGNEPNSFLKK 14
| : : : : :
Db 357 WRRGHEPQSFIIR 369

RESULT 45
TSYL1 HUMAN STANDARD; PRT; 438 AA.
ID TSYL1 HUMAN
AC Q9H0U9; O75885;
DT 15-AUG-2003, integrated into UniProtKB/Swiss-Prot.
DT 15-AUG-2003, sequence version 2.
DE Testis-specific Y-encoded-like protein 1 (TSPLY-like 1).
GN Name=TSPLY1; Synonyms=TSPLY;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Fetal brain;
RX MEDLINE=21154917; PubMed=11230166; DOI=10.1101/gr.154701;
RX Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RX Anorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RX Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RX Mews H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RX Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

```

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA] OF 366-438, AND TISSUE SPECIFICITY.
RX MEDLINE=98399864; PubMed=9730615;
RA Vogel T., Dittich O., Mehraein Y., Dechend F., Schnieders F.,
RA Schmidtke J.;
RT "Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1
RT family.";
RL Cytogenet. Cell Genet. 81:265-270(1998).
RN [4]
RP SUBCELLULAR LOCATION, AND IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=22317277; PubMed=12429849; DOI=10.1091/mbc.E02-05-0271;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
RN [5]
RP INVOLVEMENT IN SIDDT.
RX PubMed=15273283; DOI=10.1073/pnas.0401194101;
RA Puffenberger E.G., Hu-Lince D., Parod J.M., Craig D.W., Dobrin S.E.,
RA Conway A.R., Donarum E.A., Strauss K.A., Duncley T., Cardenas J.F.,
RA Melmed K.R., Wright C.A., Liang W., Stafford P., Flynn C.R.,
RA Morton D.H., Stephan D.A.;
RT "Mapping of sudden infant death with dysgenesis of the testes syndrome
RT (SIDDT) by a SNP genome scan and identification of TSPYL loss of
RT function.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:11689-11694(2004).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- TISSUE SPECIFICITY: Expressed in testis, ovary, liver, spleen,
CC brain, kidney, prostate, lung, liver, and heart.
CC -!- DISEASE: Defects in TSPYL1 are the cause of sudden infant death
CC with dysgenesis of the testes syndrome (SIDDT) [MIM:608800]. SIDDT
CC is an autosomal recessive disorder. Affected infants appear normal
CC at birth, develop signs of viscerotonic dysfunction early in
CC life, and die before 12 months of age of abrupt cardiorespiratory
CC arrest. Features included bradycardia, hypothermia, severe
CC gastroesophageal reflux, laryngospasm, bronchospasm, and abnormal
CC cardiorespiratory patterns during sleep. Genotypic males with
CC SIDDT had fetal testicular dysgenesis and ambiguous genitalia,
CC with findings such as intraabdominal testes, dysplastic testes,
CC deficient fetal testosterone production, fusion and rugation of
CC the gonadal sac, and partial development of the penile shaft.
CC Female sexual development was normal. Affected infants had an
CC unusual staccato cry, similar to the cry of a goat.
CC -!- SIMILARITY: Belongs to the nucleosome assembly protein (NAP)
CC family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AL136629; CAB66564.1; -; mRNA.
DR EMBL; BC048969; AAH48969.1; -; mRNA.
DR EMBL; AF042181; AAC62384.1; -; mRNA.
DR SWISS-2DPAGE; Q9H0U9; HUMAN.
DR Ensembl; ENSG00000189241; Homo sapiens.

```

```
DR HGNC; HGNC:12382; TSPYL1.
DR MIM; 604714; gene.
DR MIM; 608800; phenotype.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Nuclear protein; Polymorphism.
FT CHAIN 1 438
FT VARIANT 62 62 P -> S (in dbSNP:3828743).
FT VARIANT 74 74 A -> P (in dbSNP:3749895).
FT SEQUENCE 438 AA; 49292 MW; 238B9FB17736116D CRC64;
Query Match 53.7%; Score 44; DB 1; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKK 14
Db |::|::|::|
358 WRRGHEPQSFIIR 370
RESULT 46
Q6FI91_HUMAN PRELIMINARY; PRT; 438 AA.
AC Q6FI91;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE TSPYL protein.
GN Name=TSPYL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR533535; CAG38566.1; -; mRNA.
DR Ensembl; ENSG0000189241; Homo sapiens.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
SQ SEQUENCE 438 AA; 49307 MW; 834D4B7B72B59650 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKK 14
Db |::|::|::|
358 WRRGHEPQSFIIR 370
RESULT 47
Q5R5G8_PONPY PRELIMINARY; PRT; 438 AA.
AC Q5R5G8;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein DKFP469F1920.
GN Name=DKFP469F1920;
Pongo pygmaeus (Orangutan).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
NCBI_TaxID=9600;
[1]
NUCLEOTIDE SEQUENCE.
The German cDNA Consortium;
Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
Osanger A., Fobo G., Han M., Wiemann S.;
Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
EMBL; CR859130; CAH91322.1; -; mRNA.
GO; GO:0005634; C:nucleus; IEA.
GO; GO:0006334; P:nucleosome assembly; IEA.
InterPro; IPR002164; NAP_family.
PANTHER; PTHR11875; NAP_family; 1.
Pfam; PF00956; NAP; 1.
Hypothetical protein.
SEQUENCE 438 AA; 49350 MW; D4B74FFC42D9E066 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKK 14
```

```
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RG The German cDNA Consortium;
RA Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR860891; CAH92998.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49292 MW; F885B1D3D455F39B CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKK 14
Db |::|::|::|
358 WRRGHEPQSFIIR 370
RESULT 48
Q5RA88_PONPY PRELIMINARY; PRT; 438 AA.
AC Q5RA88;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein DKFP459D1728.
GN Name=DKFP459D1728;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cortex;
RG The German cDNA Consortium;
RA Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR859130; CAH91322.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49350 MW; D4B74FFC42D9E066 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNEPNSFLKK 14
```

```
Db      | :||| :|:::
358 WRRGHEPQSFIRR 370

RESULT 49
Q5QNQ4 MOUSE PRELIMINARY; PRT; 497 AA.
AC Q5QNQ4
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2006, sequence version 1.
DE Oxyesterol-binding protein.
GN Name=Osbp2; ORFNames=RP23-309E11.7-003;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Murioidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bird C.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Dunn M.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL691413; CAI25059.1; -; Genomic DNA.
DR EMBL; AL731853; CAI25059.1; JOINED; Genomic DNA.
DR EMBL; AL731853; CAI51855.1; -; Genomic DNA.
DR EMBL; AL691413; CAI51855.1; JOINED; Genomic DNA.
DR Ensembl; ENSMUSG0000020435; Mus musculus.
DR GO; GO:0006869; P:lipid transport; IEA.
DR GO; GO:0008202; P:steroid metabolism; IEA.
DR InterPro; IPR000648; Oxyesterol_bd.
DR PANTHER; PTHR10972; Oxyesterol_BP; 1.
DR Pfam; PF01237; Oxyesterol_BP; 1.
DR PROSITE; PS01013; OSBP; 1.
KW Lipid transport; Transport.
SQ SEQUENCE 497 AA; 57076 MW; 1BE492F0B494223D CRC64;

Query Match 53.7%; Score 44; DB 2; Length 497;
Best Local Similarity 72.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      2 WELGNEPNSFL 12
      ||||| :|
Db      14 WELGKPGSFL 24

RESULT 50
Q47BW1 DECAR PRELIMINARY; PRT; 492 AA.
AC Q47BW1
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2006, sequence version 1.
DE Conserved hypothetical transmembrane protein.
GN OrderedLocusNames=Daro_2940;
OS Dechloromonas aromatica (strain RCB).
OC Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales;
OC Rhodocyclaceae; Dechloromonas.
OX NCBI_TaxID=159087;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Ibrani S., Pittluck S., Di Bartolo G., Trong S., Kellar K.,
RA Schmutz J., Larimer F., Land M., Ivanova N., Richardson P.;
RT "Complete sequence of Dechloromonas aromatica RCB.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DDBJ databases.

-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000089; AAZ47670.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0020037; F:heme binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR009056; Cyt_c_monohaem.
DR InterPro; IPR012282; Cytochrome_c_R.
DR PROSITE; PS51007; CYTC; 1.
KW Complete proteome; Hypothetical protein; Transmembrane.
SQ SEQUENCE 492 AA; 54756 MW; B4C3D21079BA880B CRC64;

Query Match 53.0%; Score 43.5; DB 2; Length 492;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 10; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy      1 SWELGNE-PNSFLKK 14
      ||||| :|
Db      97 SWELGQELFIGFTKK 111

Search completed: June 5, 2006, 12:53:21
Job time : 172.507 secs
```

GenCore version 5.1.1-9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 132.219 Seconds

(without alignments)

65.702 Million cell updates/sec

Title: US-10-645-659A-7

Perfect score: 104

Sequence: 1 PAYLRFGGTKDTDFLIFDPK 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: A_Geneseq_8.*

2: Geneseqp1980s.*

3: Geneseqp1990s.*

4: Geneseqp2000s.*

5: Geneseqp2001s.*

6: Geneseqp2002s.*

7: Geneseqp2003as.*

8: Geneseqp2003bs.*

9: Geneseqp2004s.*

10: Geneseqp2005s.*

11: Geneseqp2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	19	8	Adt88213 Human hep
2	104	100.0	19	8	Adt78180 Functiona
3	104	100.0	19	8	Aea42429 Human hep
4	104	100.0	74	8	Adt88217 Human mat
5	104	100.0	74	8	Adt78184 8kDa subu
6	104	100.0	74	8	Adt7059 Heparanas
7	104	100.0	74	8	Adt18994 Human hep
8	104	100.0	74	8	Aea42433 Human mat
9	104	100.0	174	9	Adt19012 Human hep
10	104	100.0	460	9	Adt27061 Heparanas
11	104	100.0	486	9	Aeb87589 Human hep
12	104	100.0	492	9	Adt18996 Hep106 co
13	104	100.0	493	9	Aeb87562 Human hep
14	104	100.0	495	9	Adt18999 Hep109 co
15	104	100.0	497	9	Aeb87587 Human hep
16	104	100.0	501	9	Adt19000 HepGS3 co
17	104	100.0	507	9	Adt19005 HepGS6 co
18	104	100.0	508	9	Adt27058 Human ina
19	104	100.0	526	9	Adt19006 Hepyalur
20	104	100.0	527	5	Abb07815 Chicken s
21	104	100.0	527	7	Abw02018 Chimeric
22	104	100.0	527	8	Ado63825 Chimeric
23	104	100.0	527	8	Ado63827 Chimeric

24	104	100.0	527	8	Ado63826
25	104	100.0	527	9	Adt219004
26	104	100.0	530	2	Aay34173
27	104	100.0	532	2	Aay17083
28	104	100.0	543	2	AAY02345
29	104	100.0	543	2	AAY17082
30	104	100.0	543	3	AAY57590
31	104	100.0	543	3	AAB08849
32	104	100.0	543	3	AAY52990
33	104	100.0	543	4	AAY97635
34	104	100.0	543	4	AAB86206
35	104	100.0	543	4	AAB88361
36	104	100.0	543	5	ABB07813
37	104	100.0	543	7	ADD18950
38	104	100.0	543	7	ADG88800
39	104	100.0	543	8	ADL16379
40	104	100.0	543	8	ADK52086
41	104	100.0	543	8	ADM48716
42	104	100.0	543	8	ADM48759
43	104	100.0	543	8	ADN05074
44	104	100.0	543	8	ADN04902
45	104	100.0	543	8	ADO63831
46	104	100.0	543	8	ADO63824
47	104	100.0	543	8	ADO63823
48	104	100.0	543	8	ADO63832
49	104	100.0	543	8	ADO63822
50	104	100.0	543	8	ADQ80372
51	104	100.0	543	8	ADR88210
52	104	100.0	543	8	ADP25079
53	104	100.0	543	8	ADT78177
54	104	100.0	543	9	ADY27036
55	104	100.0	543	9	ADY63087
56	104	100.0	543	9	AEA42466
57	104	100.0	543	9	AEA42426
58	104	100.0	543	10	AEE96848
59	104	100.0	545	6	ABP56822
60	104	100.0	545	7	ADE16012
61	104	100.0	545	8	ADL93951
62	104	100.0	556	9	ADL219010
63	104	100.0	570	9	ADZ19008
64	104	100.0	588	2	AAY30124
65	104	100.0	592	2	AAY02346
66	104	100.0	592	3	AAB08850
67	104	100.0	592	7	ADG88804
68	104	100.0	592	8	ADL16383
69	104	100.0	592	8	ADM48720
70	104	100.0	592	9	AEA42461
71	99	95.2	22	9	ADY27042
72	99	95.2	170	8	ADL16424
73	99	95.2	170	8	ADM48758
74	99	95.2	174	9	ADZ19013
75	99	95.2	535	3	AAB08851
76	99	95.2	535	5	ABB07811
77	99	95.2	535	7	ADG88834
78	99	95.2	535	8	ADL16413
79	99	95.2	535	8	ADM48750
80	99	95.2	535	8	ADR88208
81	99	95.2	535	8	ADT78175
82	99	95.2	535	9	ADY27033
83	99	95.2	535	9	AEA42424
84	99	95.2	536	5	ABB07812
85	99	95.2	536	8	ADR88209
86	99	95.2	536	8	ADT78176
87	99	95.2	536	9	ADY27035
88	99	95.2	536	9	AEA42425
89	92	88.5	174	9	ADZ19014
90	92	88.5	545	9	ADY27034
91	81	77.9	15	9	ADU70836
92	81	77.9	15	9	ADU71279
93	81	77.9	173	9	ADZ19015
94	81	77.9	523	5	ABB07814
95	81	77.9	523	7	ABW02017
96	81	77.9	523	8	ADR88211

97	81	77.9	523	8	ADT78178	Adt78178	Chicken h	170	43	41.3	350	8	ADJ34798	Adj34798	Xylanase
98	81	77.9	523	9	ADY27037	Ady27037	Chicken h	171	43	41.3	1128	3	AAG29155	Aag29155	Arabidops
99	81	77.9	523	9	AEA24247	Aea24247	Chicken h	172	43	41.3	1206	3	AAG29154	Aag29154	Arabidops
100	80	76.9	15	9	ADU70951	Adu70951	Human hep	173	43	41.3	1293	3	AAG29153	Aag29153	Arabidops
101	77	74.0	15	9	ADU71167	Adu71167	Human hep	174	42.5	40.9	227	8	ADX66278	Adx66278	Plant ful
102	76	73.1	15	9	ADU70835	Adu70835	Human hep	175	42.5	40.9	314	8	ADX66660	Adx66660	Plant ful
103	72	69.2	15	9	ADU71278	Adu71278	Human hep	176	42.5	40.9	690	8	ADX72696	Adx72696	Plant ful
104	71	68.3	15	9	ADU71168	Adu71168	Human hep	177	42.5	40.9	770	2	AA434378	Aay434378	Amino aci
105	66	63.5	15	9	ADU71012	Adu71012	Human hep	178	42	40.4	206	6	ABP80884	Abp80884	N. gonorr
106	66	63.5	32	2	AA434175	Aay34175	Human pre	179	42	40.4	241	9	AEA22109	Aea22109	Campyloba
107	66	63.5	189	9	ADZ19016	Adz19016	Heparanas	180	42	40.4	305	3	AAG33465	Aag33465	Arabidops
108	66	63.5	282	4	AAW24147	Aam24147	Human EST	181	42	40.4	326	3	AAG33464	Aag33464	Arabidops
109	66	63.5	439	4	AAU07423	Aau07423	Human hep	182	42	40.4	404	8	ADV88147	Adv88147	Streptoco
110	66	63.5	480	4	AA497634	Aay97634	Human hep	183	42	40.4	404	8	ADV79400	Adv79400	Streptoco
111	66	63.5	480	4	AAU07418	Aau07418	Novel hum	184	42	40.4	418	3	AAG33463	Aag33463	Arabidops
112	66	63.5	480	4	AA485217	Aab85217	Heparanas	185	42	40.4	428	5	ABN93810	Abn93810	Herbicida
113	66	63.5	482	4	AA484664	Aab84664	Amino aci	186	42	40.4	429	8	ABT55647	Abt55647	Plant pol
114	66	63.5	534	4	AA485216	Aab85216	Heparanas	187	42	40.4	448	8	ABO58787	Abos8787	Human gen
115	66	63.5	534	5	ABP69310	Abp69310	Human pol	188	42	40.4	505	5	ABP30486	Abp30486	Streptoco
116	66	63.5	534	5	AA450337	Aam50337	Human pre	189	42	40.4	505	8	ABV81576	Abv81576	Streptoco
117	66	63.5	538	4	AA497633	Aay97633	Human hep	190	42	40.4	518	5	ABP27721	Abp27721	Streptoco
118	66	63.5	592	4	AA497632	Aay97632	Human hep	191	42	40.4	547	5	ABP41519	Abp41519	Human ova
119	66	63.5	592	4	AA481062	Aab81062	Human Hep	192	42	40.4	742	8	ADS10736	Adsi10736	Human the
120	66	63.5	592	4	AAU07424	Aau07424	Human hep	193	42	40.4	772	8	ADP23017	Adp23017	PRO polyp
121	66	63.5	592	4	AA485215	Aab85215	Heparanas	194	42	40.4	772	8	ADY17077	Ady17077	PRO polyp
122	65	62.5	15	9	ADU71013	Adu71013	Human hep	195	42	40.4	791	9	AED74134	Aed74134	Human pla
123	61	58.7	470	5	AA4818328	Aae18328	Human hep	196	42	40.4	856	2	AA496949	Aar96949	Phosphoen
124	61	58.7	528	5	AA4818327	Aae18327	Human hep	197	41.5	39.9	111	9	AED02424	Aed02424	T84.66 an
125	61	58.7	528	5	AA4818326	Aae18326	Human hep	198	41.5	39.9	127	8	ABO60044	Aboc60044	Human gen
126	60	57.7	15	9	ADU71166	Adu71166	Human hep	199	41.5	39.9	132	2	AAR10920	Aari10920	kappa lig
127	59	56.7	15	9	ADU71014	Adu71014	Human hep	200	41.5	39.9	145	5	ABN98792	Abn98792	Human rib
128	59	56.7	98	5	ABP10273	Abp10273	Human ORF	201	41.5	39.9	179	8	ADS17745	Adsi17745	Anti-CEA
129	55	52.9	15	9	ADU70950	Adu70950	Human hep	202	41.5	39.9	184	4	AA438805	Aam38805	Human tra
130	54	51.9	15	9	ADU70913	Adu70913	Human hep	203	41.5	39.9	184	4	AA460105	Aab60105	Human pol
131	50	48.1	9	9	ADU70628	Adu70628	Human hep	204	41.5	39.9	184	8	ADN05540	Adno5540	Antipsori
132	50	48.1	9	9	ADU70782	Adu70782	Human hep	205	41.5	39.9	184	8	ADO20501	Ado20501	Human PRO
133	50	48.1	147	9	ADZ19018	Adz19018	Heparanas	206	41.5	39.9	184	8	ADP55889	Adp55889	Human PRO
134	50	48.1	488	4	AA481469	Aab31469	Amino aci	207	41.5	39.9	184	9	ADY19167	Ady19167	PRO polyp
135	50	48.1	488	4	AA481472	Aab31472	Amino aci	208	41.5	39.9	196	4	AA440591	Aam40591	Human pol
136	50	48.1	488	4	AA481470	Aab31470	Amino aci	209	41.5	39.9	262	9	AED84408	Aed84408	Murine mo
137	49	47.1	9	2	AA434176	Aay34176	Human pre	210	41.5	39.9	623	9	AED84409	Aed84409	CAB molec
138	49	47.1	9	9	ADU70440	Adu70440	Human hep	211	41.5	39.9	710	8	AD425585	Adsa2585	Bacterial
139	49	47.1	9	9	ADU70571	Adu70571	Human hep	212	41.5	39.9	1040	8	ADS17749	Adsi17749	Rat major
140	49	47.1	9	9	ADU70351	Adu70351	Human hep	213	41.5	39.9	1072	8	ADS17747	Adsi17747	Human maj
141	49	47.1	9	9	ADU70811	Adu70811	Human hep	214	41.5	39.9	1127	8	ADS17789	Adsi17789	Polylysin
142	49	47.1	15	9	ADU70856	Adu70856	Human hep	215	41.5	39.9	1159	8	ADS17787	Adsi17787	Polylysin
143	48	46.2	9	9	ADU70535	Adu70535	Human hep	216	41.5	39.9	1206	8	ADS17763	Adsi17763	GAL4 + ra
144	48	46.2	9	9	ADU70629	Adu70629	Human hep	217	41.5	39.9	1238	8	ADS17761	Adsi17761	GAL4 + hu
145	48	46.2	151	9	ADZ19017	Adz19017	Heparanas	218	41.5	39.9	1273	8	ADS17777	Adsi17777	MS2 pepti
146	48	46.2	261	4	AA4859593	Aab59593	Human car	219	41	39.4	116	8	ADL05804	Adl05804	M. catarr
147	48	46.2	265	5	ABN90640	Abn90640	Chlamydia	220	41	39.4	138	9	ADZ19021	Adz19021	Heparanas
148	48	46.2	279	2	AA434803	Aay34803	Chlamydia	221	41	39.4	274	8	ADX88300	Adx88300	Plant ful
149	48	46.2	411	5	ABP66200	Abp66200	Bifidobac	222	41	39.4	274	8	ADY12676	Ady12676	Plant ful
150	47	45.2	9	9	ADU70684	Adu70684	Human hep	223	41	39.4	278	8	ADX80584	Adx80584	Plant ful
151	47	45.2	9	9	ADU70594	Adu70594	Human hep	224	41	39.4	281	8	ADX66918	Adx66918	Plant ful
152	47	45.2	9	9	ADU70685	Adu70685	Human hep	225	41	39.4	333	5	ABP27264	Abp27264	Streptoco
153	46	44.2	150	9	ADZ19019	Adz19019	Heparanas	226	41	39.4	342	3	AAG14301	Aag14301	Arabidops
154	46	44.2	414	4	ABG25051	Abg25051	Novel hum	227	41	39.4	346	3	AAG14300	Aag14300	Arabidops
155	46	44.2	488	4	AA431471	Aab31471	Amino aci	228	41	39.4	351	6	ABM70737	Abm70737	Photorhab
156	45	43.3	9	9	ADU70683	Adu70683	Human hep	229	41	39.4	367	8	ADT58411	Adt58411	Plant pol
157	45	43.3	486	8	ADI45111	Adi45111	Rice isop	230	41	39.4	412	3	AAG14299	Aag14299	Arabidops
158	45	43.3	3540	8	ADN96836	Adn96836	Bugula br	231	41	39.4	442	8	ADX66334	Adx66334	Plant ful
159	44	42.3	9	9	ADU70686	Adu70686	Human hep	232	41	39.4	694	9	ABM91425	Abm91425	M. xanthu
160	44	42.3	15	9	ADU70952	Adu70952	Human hep	233	41	39.4	754	3	AB35987	Ab35987	Sorbitol
161	44	42.3	87	4	AAU55506	Aau55506	Propionib	234	41	39.4	848	2	AA434512	Aay34512	Porphorym
162	44	42.3	87	6	ABM52025	Abm52025	Propionib	235	41	39.4	878	2	AA434385	Aay34385	Porphorym
163	44	42.3	383	8	ADX90305	Adx90305	Plant ful	236	41	39.4	5069	2	AAW52846	Aaw52846	A. medite
164	44	42.3	569	8	ADY23112	Ady23112	Plant ful	237	40	38.5	76	7	ABO72337	AbO72337	Pseudomon
165	44	42.3	659	1	AA492106	Aap92106	Bacillus	238	40	38.5	96	3	AB41008	Aab41008	Human ORF
166	44	42.3	779	5	ABM07845	Abm07845	Human MAP	239	40	38.5	96	5	ABP34610	Abp34610	Human ORF
167	44	42.3	997	8	ADX66509	Adx66509	Plant ful	240	40	38.5	113	4	AAO06297	Aao06297	Human pol
168	44	42.3	998	9	ADX05670	Adx05670	Cyclin-de	241	40	38.5	158	6	ADA34928	Ada34928	Acinetoba
169	43	41.3	253	6	ABU48538	Abu48538	Protein e	242	40	38.5	178	5	ABB48544	Abb48544	Listeria

243	40	38.5	178	6	ABU32420	Abu32420 Protein e	316	39	37.5	113	4	AAU31556	Aau31556 Novel hum
244	40	38.5	204	4	AAU19298	Aau19298 Human G p	317	39	37.5	119	4	AAO03948	Aao03948 Human pol
245	40	38.5	211	4	AGS90850	Aag90850 C glutami	318	39	37.5	127	2	AAI12885	Aay12885 Human 5'
246	40	38.5	240	3	AAS57080	Abb57080 Human pro	319	39	37.5	149	4	ABG23456	Abg23456 Novel hum
247	40	38.5	262	4	ABB66342	Abb66342 Drosophil	320	39	37.5	150	5	ABP07825	Abp07825 Human ORF
248	40	38.5	287	4	ABB66343	Abb66343 Drosophil	321	39	37.5	158	5	ABM70650	Abm70650 Photorhab
249	40	38.5	297	3	AAI15814	Aagi15814 Arabidops	322	39	37.5	225	6	ABM70650	Abm70650 Photorhab
250	40	38.5	299	7	ADP40114	Adp40114 Human NOV	323	39	37.5	227	3	AAO07227	Aao07227 Arabidops
251	40	38.5	302	4	ABB69502	Abb69502 Drosophil	324	39	37.5	235	4	AAO07227	Aao07227 Arabidops
252	40	38.5	326	4	ABB71638	Abb71638 Drosophil	325	39	37.5	235	4	AAO07227	Aao07227 Arabidops
253	40	38.5	331	3	AAI95072	Aay95072 Candida a	326	39	37.5	241	8	ADX71344	Adx71344 Plant ful
254	40	38.5	333	8	ADV88472	Adv88472 Streptoco	327	39	37.5	279	7	ADE12789	Adel12789 L. rhamno
255	40	38.5	333	8	ADV79725	Adv79725 Streptoco	328	39	37.5	281	3	AAO07226	Aao07226 Arabidops
256	40	38.5	333	8	ADV81883	Adv81883 Streptoco	329	39	37.5	281	3	ADP88603	Adp88603 Ribosome
257	40	38.5	356	6	ABG73525	Abg73525 B. subtil	330	39	37.5	284	9	ADV34638	Adv34638 Des-val a
258	40	38.5	362	8	ADP01757	Adp01757 Human sel	331	39	37.5	298	4	ABG06306	Abg06306 Novel hum
259	40	38.5	362	8	ADP90527	Adp90527 Human sel	332	39	37.5	298	7	ADF60292	Adf60292 Human con
260	40	38.5	364	8	ADX77474	Adx77474 Plant ful	333	39	37.5	298	7	ADF58916	Adf58916 Human pol
261	40	38.5	365	2	AAI47874	Aai47874 Enzyme/bi	334	39	37.5	298	7	ADF60477	Adf60477 Human con
262	40	38.5	365	2	AAW09395	Aaw09395 Rhodococc	335	39	37.5	305	8	ADN47029	Adn47029 Thermococ
263	40	38.5	365	2	AAW97239	Aaw97239 An enzyme	336	39	37.5	313	8	ADY07646	Ady07646 Plant ful
264	40	38.5	365	2	AAW89242	Aaw89242 Rhodococc	337	39	37.5	313	8	ADX93718	Adx93718 Plant ful
265	40	38.5	365	4	AAU02430	Aau02430 Rhodococc	338	39	37.5	319	3	AAO07225	Aao07225 Arabidops
266	40	38.5	365	7	AAE37615	Aae37615 Rhodococc	339	39	37.5	326	7	ABO75715	Ab075715 Pseudomon
267	40	38.5	373	3	AAI15813	Aagi15813 Arabidops	340	39	37.5	338	4	ABG23449	Abg23449 Novel hum
268	40	38.5	380	7	ADN95151	Adn95151 Human BEC	341	39	37.5	340	4	ABP99240	Abp99240 Microsom
269	40	38.5	381	3	AAO03188	Aao03188 Human sel	342	39	37.5	342	6	ABP99240	Abp99240 Microsom
270	40	38.5	381	5	AAU84306	Aau84306 Human end	343	39	37.5	342	7	ABO76674	Ab076674 Pseudomon
271	40	38.5	381	6	ABU89753	Abu89753 Protein d	344	39	37.5	347	8	ADX70642	Adx70642 Plant ful
272	40	38.5	381	7	ADD47226	Add47226 Human Pro	345	39	37.5	355	8	ADT57590	Adt57590 Plant pol
273	40	38.5	381	7	ADP65180	Adp65180 Human sel	346	39	37.5	355	8	ADT57590	Adt57590 Plant pol
274	40	38.5	381	7	ABE86458	Abe86458 Human sel	347	39	37.5	362	6	AAO16382	Aao16382 Drosophil
275	40	38.5	381	7	ADU69218	Adu69218 Human sel	348	39	37.5	373	8	ADS27581	Ads27581 Bacterial
276	40	38.5	381	9	AEC12741	Aec12741 Human sur	349	39	37.5	376	8	ADY09636	Ady09636 Plant ful
277	40	38.5	381	9	AEC12381	Aec12381 Human sur	350	39	37.5	385	9	ADD47224	Add47224 Rat Prote
278	40	38.5	381	10	ABE88077	Abe88077 Human pro	351	39	37.5	385	9	AEC12594	Aec12594 Rat surro
279	40	38.5	404	3	AAI15812	Aai15812 Arabidops	352	39	37.5	385	9	AEC12594	Aec12594 Rat surro
280	40	38.5	404	8	ADT55977	Adt55977 Plant pol	353	39	37.5	387	4	ABU52973	Abu52973 Human tes
281	40	38.5	477	8	ADS43285	Ads43285 Bacterial	354	39	37.5	391	5	ABP62776	Abp62776 Protein f
282	40	38.5	512	9	ABM92453	Abm92453 M. xanthu	355	39	37.5	391	7	ADJ72187	Adj72187 S roseosp
283	40	38.5	546	5	ABG93260	Abg93260 C. albica	356	39	37.5	394	4	AAG89139	Aag89139 Human sec
284	40	38.5	546	5	ABG93297	Abg93297 C. albica	357	39	37.5	394	7	ADM05625	Adm05625 Human pro
285	40	38.5	550	5	ABG93259	Abg93259 C. albica	358	39	37.5	394	7	ADM05625	Adm05625 Human pro
286	40	38.5	550	5	ABG93311	Abg93311 C. albica	359	39	37.5	404	9	AEC55897	Aec55897 Saccharom
287	40	38.5	550	5	ABG93365	Abg93365 Human BAX	360	39	37.5	417	4	ABB71426	Abb71426 Drosophil
288	40	38.5	580	4	AAI90361	Aai90361 C glutami	361	39	37.5	419	4	ABB66118	Abb66118 Drosophil
289	40	38.5	585	8	ADN22546	Adn22546 Bacterial	362	39	37.5	423	4	ABB60949	Abb60949 Drosophil
290	40	38.5	633	4	AAI75283	Aai75283 Corynebact	363	39	37.5	426	8	ABE86882	Aeb86882 Polyketid
291	40	38.5	637	4	ABG08572	Abg08572 Novel hum	364	39	37.5	426	8	ABE86679	Aeb86679 Polyketid
292	40	38.5	637	4	ABG03188	Abg03188 Novel hum	365	39	37.5	441	7	ADG28432	Adg28432 Soybean S
293	40	38.5	701	8	ADS03240	Ads03240 Bacterial	366	39	37.5	450	7	ADG28432	Adg28432 Soybean S
294	40	38.5	713	9	ABM97151	Abm97151 M. xanthu	367	39	37.5	539	8	ADY13296	Ady13296 Plant ful
295	40	38.5	742	2	AAO05235	Aao05235 Amino aci	368	39	37.5	561	5	ABG97480	Abg97480 Human NOV
296	40	38.5	743	4	ABG23278	Abg23278 Novel hum	369	39	37.5	651	3	AAI10936	Aai10936 Human RNA
297	40	38.5	743	4	ABG14502	Abg14502 Novel hum	370	39	37.5	651	3	ABE78308	Abe78308 Amino aci
298	40	38.5	7968	10	ABE68641	Abe68641 Streptomy	371	39	37.5	661	8	ADN24139	Adn24139 Bacterial
299	39.5	38.0	388	7	AAI32058	Aai32058 Cat pregn	372	39	37.5	730	9	AEA49153	Aea49153 L. rhamno
300	39.5	38.0	388	7	ADP11571	Adp11571 Feline pr	373	39	37.5	730	9	AEA49153	Aea49153 L. rhamno
301	39.5	38.0	551	8	ADS23997	Ads23997 Bacterial	374	39	37.5	735	7	AEM86341	Aem86341 Rice abio
302	39.5	38.0	584	8	ADL04704	Adl04704 M. catar	375	39	37.5	756	8	ADX91671	Adx91671 Plant ful
303	39	37.5	9	9	ADU70627	Adu70627 Human hep	376	39	37.5	782	7	ADP93177	Adp93177 Human N-a
304	39	37.5	15	9	ADU10111	Adu10111 Human hep	377	39	37.5	782	7	ADP93183	Adp93183 Human N-a
305	39	37.5	31	4	AAI18407	Aai18407 Peptide #	378	39	37.5	784	7	ADP93185	Adp93185 Human N-a
306	39	37.5	31	4	ABB37446	Abb37446 Peptide #	379	39	37.5	791	7	ADP93179	Adp93179 Murine N-
307	39	37.5	31	4	AAI30883	Aai30883 Peptide #	380	39	37.5	792	8	ADO28695	Ado28695 Human can
308	39	37.5	31	4	ABB32197	Abb32197 Peptide #	381	39	37.5	792	8	ADS32728	Ads32728 Human N-a
309	39	37.5	31	4	ABB22742	Abb22742 Protein #	382	39	37.5	792	8	ADU05159	Adu05159 Human N-a
310	39	37.5	31	4	AAI70567	Aai70567 Human bon	383	39	37.5	821	5	AAE25023	Aae25023 Human dru
311	39	37.5	31	4	AAI58124	Aai58124 Human bra	384	39	37.5	875	8	ADS28775	Ads28775 Bacterial
312	39	37.5	31	4	ABG52248	Abg52248 Human liv	385	39	37.5	909	3	AAI94328	Aai94328 Maize DNA
313	39	37.5	31	4	AAI06007	Aai06007 Peptide #	386	39	37.5	1360	8	ADO26836	Ado26836 Human rec
314	39	37.5	31	5	ABG40231	Abg40231 Human pep	387	39	37.5	5432	8	ADQ91708	Adq91708 Polyketid
315	39	37.5	67	4	ABG23453	Abg23453 Novel hum	388	39	37.5	5432	8	ABE87000	Aeb87000 Streptomy

389	39	37.5	5432	8	ABE86797	Ab86797 Streptomy	462	38	36.5	503	7	ABO69145	Ab069145 Pseudom
390	38.5	37.0	67	6	ADA32924	Ada32924 Acinetoba	463	38	36.5	525	6	ADA56830	Ada56830 Human sec
391	38.5	37.0	120	3	AAY56701	Aay56701 Rat anti-	464	38	36.5	525	6	ABR47695	AbR47695 Human sec
392	38.5	37.0	186	6	ABM87982	Abm87982 Rice abio	465	38	36.5	525	6	ABR00046	AbR00046 Human gen
393	38.5	37.0	322	6	ADB17492	Adb17492 Soybean p	466	38	36.5	525	7	ADB91501	AdB91501 Human sec
394	38.5	37.0	322	9	AEC75716	Aec75716 Soybean A	467	38	36.5	525	7	ADC74074	AdC74074 Human sec
395	38.5	37.0	397	8	ADN46729	Adn46729 Thermococ	468	38	36.5	530	4	ABBS7868	AbBS7868 Drosophil
396	38.5	37.0	408	3	AAY87978	Aay87978 Pseudomon	469	38	36.5	540	8	ADQ89648	AdQ89648 Antagonis
397	38.5	37.0	539	8	ADN17319	D. destru	470	38	36.5	553	2	ADN20097	AdN20097 Bacterial
398	38.5	37.0	574	3	AAY87977	Aay87977 Pseudomon	471	38	36.5	573	2	AAW48874	Aaw48874 Hyphozyma
399	38.5	37.0	667	3	AAY87976	Aay87976 Pseudomon	472	38	36.5	577	6	ABU32269	Abu32269 Protein e
400	38.5	37.0	984	5	ABBS9326	Abbs9326 Herbicida	473	38	36.5	582	7	ABO66203	AbO66203 Klebsiell
401	38.5	37.0	1545	9	ADZ44910	Adz44910 D. carota	474	38	36.5	599	6	ADA33364	Ada33364 Acinetoba
402	38	36.5	9	9	ADU70466	Adu70466 Human hep	475	38	36.5	602	8	AAW47598	Aaw47598 Drosophil
403	38	36.5	15	9	ADU71269	Adu71269 Human hep	476	38	36.5	625	8	ADS42094	AdS42094 Bacterial
404	38	36.5	20	2	AAY38490	Aay38490 Human sec	477	38	36.5	738	2	AAR13993	Aar13993 A. altocet
405	38	36.5	45	4	ABBA1509	Abba1509 Peptide #	478	38	36.5	738	2	AAR20192	Aar20192 ADH compl
406	38	36.5	45	4	AAM35301	Aam35301 Peptide #	479	38	36.5	761	4	AAB46719	Aab46719 T. brucei
407	38	36.5	45	4	ABB25382	Abb25382 Protein #	480	38	36.5	785	9	ADL98278	AdL98278 Human ins
408	38	36.5	45	4	AAM75189	Aam75189 Human bon	481	38	36.5	785	9	ADW00634	AdW00634 Human pro
409	38	36.5	45	4	AAM62379	Aam62379 Human bra	482	38	36.5	834	4	AAB61238	Aab61238 Murine M-
410	38	36.5	45	4	ABGS6951	Abgs6951 Human liv	483	38	36.5	834	6	ABO32675	AbO32675 Secreted
411	38	36.5	45	5	ABG44852	Abg44852 Human pep	484	38	36.5	834	7	ABD90780	AbD90780 Mouse M-S
412	38	36.5	78	4	ABB17364	Abb17364 Human ner	485	38	36.5	834	7	ADF71515	AdF71515 Murine M-
413	38	36.5	131	2	AAW30278	Aaw30278 Light cha	486	38	36.5	834	8	ADQ10333	AdQ10333 Human pol
414	38	36.5	138	9	ADZ19020	Adz19020 Heparanas	487	38	36.5	834	8	ADR67091	AdR67091 Mouse can
415	38	36.5	170	4	AG74676	Ag74676 Human col	488	38	36.5	856	6	ADA55361	Ada55361 Human pro
416	38	36.5	217	8	ADS27809	Ads27809 Bacterial	489	38	36.5	883	8	ADI45397	AdI45397 Rice isop
417	38	36.5	225	3	AAG13478	Ag13478 Arabidops	490	38	36.5	896	8	ADX76298	AdX76298 Plant ful
418	38	36.5	238	9	ABM96838	Abm96838 M. xanthu	491	38	36.5	921	9	ADD46793	AdD46793 Human pro
419	38	36.5	240	9	ABM96678	Abm96678 M. xanthu	492	38	36.5	921	9	AEF53848	Aef53848 Human ins
420	38	36.5	247	3	AAG13477	Ag13477 Arabidops	493	38	36.5	924	8	ABBS4116	AbBS4116 Lactococ
421	38	36.5	253	3	AGA33712	Ag33712 Arabidops	494	38	36.5	924	8	ADS29350	AdS29350 Bacterial
422	38	36.5	255	8	ADS21205	Ads21205 Bacterial	495	38	36.5	1058	6	ABU26274	AbU26274 Protein e
423	38	36.5	257	3	AAG43711	Ag43711 Arabidops	496	38	36.5	1079	5	ABP73500	AbP73500 Candida a
424	38	36.5	273	7	ABR62846	AbR62846 Human car	497	38	36.5	1141	7	ABM89610	Abm89610 Rice abio
425	38	36.5	274	2	AAY38486	Aay38486 Human sec	498	38	36.5	1157	8	ADP99144	AdP99144 Human tra
426	38	36.5	276	8	ADX78137	Adx78137 Plant ful	499	38	36.5	1213	8	ADP98964	AdP98964 C. albica
427	38	36.5	278	6	ABU43175	Abu43175 Protein e	500	38	36.5	1268	4	ABB11702	AbB11702 Human ins
428	38	36.5	279	3	AGA33710	Ag33710 Arabidops	501	38	36.5	1268	4	ABE11702	AbE11702 Human ins
429	38	36.5	279	5	ABP39377	Abp39377 Staphyloc	502	38	36.5	1268	7	ADE09341	Ade09341 Novel pro
430	38	36.5	279	8	ADS04554	AdS04554 Staphyloc	503	38	36.5	1282	5	AAE19143	Aae19143 Human kin
431	38	36.5	284	6	ABU21840	Abu21840 Protein e	504	38	36.5	1297	3	AAAB36840	AaB36840 Human ins
432	38	36.5	311	8	ADN22890	Adn22890 Bacterial	505	38	36.5	1297	9	AEF53847	Aef53847 Human ins
433	38	36.5	341	8	ADN22896	Adn22896 Bacterial	506	38	36.5	1310	6	ABP96077	AbP96077 Human pro
434	38	36.5	343	8	ADX71554	Adx71554 Plant ful	507	38	36.5	1374	8	ADN22758	AdN22758 Bacterial
435	38	36.5	353	7	ADD48962	AdD48962 Rat Prote	508	38	36.5	1400	7	ADL33323	AdL33323 Human tra
436	38	36.5	353	7	ADD48170	AdD48170 Rat Prote	509	38	36.5	1418	4	ABG02184	AbG02184 Novel hum
437	38	36.5	359	2	AAR97868	Aar97868 Hamster p	510	38	36.5	1435	10	AEF72192	Aef72192 Human tar
438	38	36.5	359	7	ADE60176	Ade60176 Human pro	511	38	36.5	1440	7	ADE08390	Ade08390 Novel pro
439	38	36.5	359	7	ADDE60174	Adde60174 Rat Prote	512	38	36.5	1469	6	ABR43185	AbR43185 Human REM
440	38	36.5	359	7	ADDA48964	AdDa48964 Human pro	513	38	36.5	1469	8	ABM84440	Abm84440 Human dia
441	38	36.5	359	7	ADDA48172	AdDa48172 Human pro	514	38	36.5	1503	3	AAAY92944	Aay92944 Human TRP
442	38	36.5	359	7	ABO66595	AbO66595 Klebsiell	515	38	36.5	1503	3	AAAY95439	Aay95439 Human cal
443	38	36.5	359	9	ADX57715	Adx57715 Rheumatoi	516	38	36.5	1503	3	AAAB36865	AaB36865 Human put
444	38	36.5	359	9	ADY15754	Ady15754 PRO polyp	517	38	36.5	1503	5	ABB76459	AbB76459 Human lon
445	38	36.5	360	6	ABP79596	AbP79596 N. gonorr	518	38	36.5	1503	5	ABBA84549	AbB84544 Human tra
446	38	36.5	361	5	ABB91031	Abb91031 Herbicida	519	38	36.5	1503	7	ADC47022	AdC47022 Human LTR
447	38	36.5	364	8	ADT58959	Adt58959 Plant pol	520	38	36.5	1503	7	ADC77685	AdC77685 Human 222
448	38	36.5	406	8	ADRO9028	Adro9028 Human pro	521	38	36.5	1503	7	ADC83633	Adc83633 LTRPC3-re
449	38	36.5	406	9	AEA19966	Aea19966 Novel hum	522	38	36.5	1503	8	ADQ89102	AdQ89102 Human uro
450	38	36.5	409	6	ABU00700	Abu00700 S. pneumo	523	38	36.5	1503	8	ADT93474	AdT93474 Human tra
451	38	36.5	409	6	ABP81316	Abp81316 Streptoco	524	38	36.5	1503	9	ABE85144	AbE85144 Human nov
452	38	36.5	409	6	ABU45740	Abu45740 Protein e	525	38	36.5	1583	5	ABP70147	AbP70147 Human NOV
453	38	36.5	414	6	ABU36068	Abu36068 Protein e	526	38	36.5	1585	5	AAE21186	Aae21186 Human TRI
454	38	36.5	414	7	ADB74341	Adb74341 Mycobacte	527	38	36.5	1619	6	ADA36060	Ada36060 Acinetoba
455	38	36.5	426	8	ABE86878	AbE86878 Polyketid	528	38	36.5	1624	5	ABB81576	AbB81576 Human ATP
456	38	36.5	426	8	ABE86675	AbE86675 Polyketid	529	38	36.5	1624	5	ABB81582	AbB81582 Human ABC
457	38	36.5	435	6	ABU40192	Abu40192 Protein e	530	38	36.5	1624	6	ABB84638	AbB84638 Human ABC
458	38	36.5	440	4	AAU33635	Aau33635 Pseudomon	531	38	36.5	1624	10	AEF72191	Aef72191 Human tar
459	38	36.5	440	6	ABU38545	Abu38545 Protein e	532	38	36.5	1808	10	AEF57853	Aef57853 Polyketid
460	38	36.5	461	6	ABU20543	Abu20543 Protein e	533	38	36.5	2088	10	AEF16945	Aef16945 Streptomy
461	38	36.5	469	7	ABM88848	Abm88848 Rice abio	534	38	36.5	5207	8	ADQ91706	Adq91706 Polyketid

535	38	36.5	5207	8	AE8B6998	Aeb86698 Streptomy	608	37	35.6	389	2	AAW53158	Aaw53158 Arthrobac
536	38	36.5	5207	8	AE8B6795	Aeb86795 Streptomy	609	37	35.6	389	2	AAW53162	Aaw53162 Arthrobac
537	38	36.5	5712	10	AE8E8639	Aee68639 Streptomy	610	37	35.6	389	2	AAW53161	Aaw53161 Arthrobac
538	38	36.5	7257	3	AA558576	Aay58576 Sorangium	611	37	35.6	389	2	AAW53164	Aaw53164 Arthrobac
539	37.5	36.1	72	3	AAG02738	Aag02738 Human sec	612	37	35.6	389	2	AAW53160	Aaw53160 Arthrobac
540	37.5	36.1	123	7	ADL96646	Adl96646 M. paratu	613	37	35.6	389	2	AAB99724	Aab99724 Arthrobac
541	37.5	36.1	227	5	ABG70073	Abg70073 Human pre	614	37	35.6	389	8	ADO23559	Ado23559 Arthrobac
542	37.5	36.1	246	7	ADC87249	Adc87249 Human GPC	615	37	35.6	390	1	AA94653	Aap94653 New sarco
543	37.5	36.1	262	9	AED84412	Aed84412 Modified	616	37	35.6	390	1	AA94653	Aap94653 New sarco
544	37.5	36.1	274	9	ADY66716	Ady66716 S. mansoni	617	37	35.6	390	1	AA94653	Aap94653 New sarco
545	37.5	36.1	347	4	ABB59052	Abb59052 Drosophil	618	37	35.6	397	5	ABP66124	Abp66124 Bifidobac
546	37.5	36.1	375	5	ABB57201	Abb57201 Mouse tcc	619	37	35.6	401	6	ABU32343	Abu32343 Protein e
547	37.5	36.1	375	9	ADK06407	Adk06407 Cyclin-de	620	37	35.6	402	7	ABO67435	Abo67435 Klebsiell
548	37.5	36.1	411	7	ADC30913	Adc30913 Human nov	621	37	35.6	403	8	ADT58194	Adt58194 Plant pol
549	37.5	36.1	430	9	ACA89042	Aca89042 Chicken t	622	37	35.6	405	7	ADC96787	Adc96787 E. faeciu
550	37.5	36.1	623	3	AEA99706	Aea99706 Human CAB	623	37	35.6	410	4	ABB65321	Abb65321 Drosophil
551	37.5	36.1	623	9	AED84415	Aed84415 CAB molec	624	37	35.6	411	7	ABG29768	Abg29768 Arthrobac
552	37.5	36.1	623	9	AED84414	Aed84414 CAB molec	625	37	35.6	418	4	ABG29768	Abg29768 Novel hum
553	37.5	36.1	1296	2	AAW88448	Aaw88448 Caenorhab	626	37	35.6	420	8	ADP45524	Adp45524 Human col
554	37.5	36.1	1559	9	ADZ44912	Adz44912 L. esculen	627	37	35.6	425	4	ABB62474	Abb62474 Drosophil
555	37	35.6	31	2	AAW62083	Aaw62083 Hyphozyma	628	37	35.6	427	7	ADF05940	Adf05940 Bacterial
556	37	35.6	75	4	AAW86624	Aaw86624 Human imm	629	37	35.6	430	6	ADA34337	Ada34337 Acinetoba
557	37	35.6	135	8	ADK66293	Adk66293 Plant ful	630	37	35.6	432	6	ABU26569	Abu26569 Protein e
558	37	35.6	138	7	ADM04818	Adm04818 Human pro	631	37	35.6	438	6	ABM15924	Abm15924 Mycobacte
559	37	35.6	138	9	AEC87748	Aec87748 Human CDN	632	37	35.6	439	6	ABU24183	Abu24183 Protein e
560	37	35.6	142	4	AAU55549	Aau55549 Propionib	633	37	35.6	442	8	ADX77733	Adx77733 Plant ful
561	37	35.6	142	6	ABM52068	Abm52068 Propionib	634	37	35.6	459	6	ABU20396	Abu20396 Protein e
562	37	35.6	149	6	ABU44605	Abu44605 Protein e	635	37	35.6	470	7	ABU80211	Abu80211 Mycobacte
563	37	35.6	151	4	AAO04698	Aao04698 Human pol	636	37	35.6	478	6	ABG74844	Abg74844 Potato 9-
564	37	35.6	211	1	AAPI0032	Aapi0032 Sequence	637	37	35.6	496	8	AQO08798	Aqo08798 Clona int
565	37	35.6	227	1	AAPI0035	Aapi0035 Sequence	638	37	35.6	505	2	AAI35699	Aay35699 C. pneumo
566	37	35.6	228	7	ADJ11584	Adj11584 Rice prot	639	37	35.6	516	5	ABY47838	Abh47838 Listeria
567	37	35.6	254	5	ADP28106	Adp28106 Streptoco	640	37	35.6	517	4	ADH87433	Adh87433 Enterococ
568	37	35.6	254	8	ADV88855	Adv88855 Streptoco	641	37	35.6	541	4	ABE62565	Abh62565 Drosophil
569	37	35.6	254	8	ADV82241	Adv82241 Streptoco	642	37	35.6	568	7	ADJ70002	Adj70002 Human hea
570	37	35.6	254	8	ADV80108	Adv80108 Streptoco	643	37	35.6	572	8	ADX73963	Adx73963 Plant ful
571	37	35.6	266	7	ADJ11794	Adj11794 Rice prot	644	37	35.6	578	8	ADN23662	Adn23662 Bacterial
572	37	35.6	266	7	ADJ11440	Adj11440 Rice prot	645	37	35.6	607	6	ABG072573	Abg072573 BSA conta
573	37	35.6	272	7	ABR42950	Abra42950 Human exo	646	37	35.6	607	6	ABO07192	Abu07192 Human p53
574	37	35.6	296	8	ADK46448	Adk46448 Streptoco	647	37	35.6	607	7	ABE63300	Ade63300 Human pro
575	37	35.6	306	7	ADF07209	Adf07209 Bacterial	648	37	35.6	607	8	AQO95952	Adq95952 T cell ac
576	37	35.6	309	7	ABO63712	Abo63712 Klebsiell	649	37	35.6	607	8	AQO95990	Adq95990 T cell ac
577	37	35.6	314	4	AA879059	Aab79059 Corynebac	650	37	35.6	638	8	ADU25696	Adu25696 L. acidop
578	37	35.6	314	4	AAG91752	Aag91752 C. glutami	651	37	35.6	638	9	AEC57426	Aec57426 L. acidop
579	37	35.6	314	7	ADD13753	Adi13753 C. glutam	652	37	35.6	684	8	ADS42109	Adsa42109 Bacterial
580	37	35.6	319	7	ADM26740	Adm26740 Hyperther	653	37	35.6	695	7	ADB70236	C. neofo
581	37	35.6	331	6	ADM25756	Adm25756 Hyperther	654	37	35.6	700	6	ABU23181	Abu23181 Protein e
582	37	35.6	336	6	ADP81638	Adp81638 Streptoco	655	37	35.6	884	6	ABR43327	Abra43327 Dictyoglo
583	37	35.6	337	4	ABB67824	Abb67824 Drosophil	656	37	35.6	884	6	ABR43348	Abra43348 Dictyoglo
584	37	35.6	337	8	ADNA46602	Adna46602 Thermococ	657	37	35.6	929	5	ABG30539	Abg30539 Alpha-iso
585	37	35.6	361	8	ADP45523	Adp45523 Human col	658	37	35.6	962	6	ABR42511	Abra42511 Clorobioc
586	37	35.6	364	6	ADA34721	Ada34721 Acinetoba	659	37	35.6	965	5	ABG30565	Abg30565 Alpha-iso
587	37	35.6	366	9	AED31464	Aed31464 Rubber al	660	37	35.6	965	6	ADA26483	Ada26483 Alpha-iso
588	37	35.6	374	5	ABB93667	Abb93667 Herbicida	661	37	35.6	990	8	ADN73117	Adn73117 Thale cre
589	37	35.6	382	5	ABG93031	Abg93031 S. cerevi	662	37	35.6	1152	6	ABM67949	Abm67949 Photorhab
590	37	35.6	382	8	ADN18703	Adn18703 Bacterial	663	37	35.6	1289	6	ABK52641	Abk52641 Protein s
591	37	35.6	387	2	AA838078	Aar38078 Sarcosine	664	37	35.6	1289	7	ABK63056	Abk63056 Disease t
592	37	35.6	387	9	ADK02873	Adk02873 Novel mod	665	37	35.6	1873	2	AA873055	Aar73055 Rabbit sk
593	37	35.6	387	9	AEB07361	Aeb07361 Novel mod	666	37	35.6	1873	2	AAW18390	Aaw18390 Rabbit ca
594	37	35.6	389	2	AA826658	Aar52658 Arthrobac	667	37	35.6	1873	2	AAW37711	Aaw37711 Rabbit sk
595	37	35.6	389	2	AA879150	Aar79150 Arthrobac	668	37	35.6	1873	3	AA877544	Aay77544 Rabbit sk
596	37	35.6	399	2	AA876735	Aar76735 N-methyl	669	37	35.6	1873	3	AA876777	Aeb76777 Rabbit sk
597	37	35.6	399	2	AAW25149	Aaw25149 Modified	670	37	35.6	1873	9	ABE26777	Aeb26777 Rabbit sk
598	37	35.6	399	2	AAW25150	Aaw25150 Modified	671	37	35.6	1925	6	AAE37000	Aae37000 Micromono
599	37	35.6	399	2	AAW25148	Aaw25148 Arthrobac	672	37	35.6	2108	10	AEF66135	Aef66135 Chick Ovo
600	37	35.6	399	2	AAW25151	Aaw25151 Modified	673	37	35.6	2108	10	AEF66132	Aef66132 Chick Ovo
601	37	35.6	399	2	AAW71461	Aaw71461 A modifie	674	37	35.6	4032	8	ADV99897	Adv99897 Nanchangm
602	37	35.6	399	2	AAW71462	Aaw71462 A modifie	675	37	35.6	6532	10	AEF57850	Aef57850 Polyketid
603	37	35.6	399	2	AAW53163	Aaw53163 Arthrobac	676	36.5	35.1	32	9	ABE21381	Aeb21381 Mouse ant
604	37	35.6	399	2	AAW53156	Aaw53156 Arthrobac	677	36.5	35.1	32	9	ABE31139	Aeb31139 Antibody
605	37	35.6	399	2	AAW53159	Aaw53159 Arthrobac	678	36.5	35.1	52	6	ABU56917	Abu56917 BONT/A Hc
606	37	35.6	399	2	AAW53157	Aaw53157 Arthrobac	679	36.5	35.1	53	6	ABU56848	Abu56848 BONT/A Hc
607	37	35.6	399	2	AAW53155	Aaw53155 Arthrobac	680	36.5	35.1	55	5	ABP09210	Abp09210 Human ORF

827	36	34.6	249	7	ADC45087	Adc45087 S. pneumo	900	36	34.6	437	5	AAO20319	Aao20319 Protein o
828	36	34.6	250	2	AAW61251	Aaw61251 Streptoco	901	36	34.6	437	5	ABG80323	Abg80323 C. glutam
829	36	34.6	250	5	ABP54669	Abp54669 S. pneumo	902	36	34.6	437	8	ADR28244	Adr28244 Corynebac
830	36	34.6	250	7	ADC45307	Adc45307 S. pneumo	903	36	34.6	437	9	AEC56498	Aec56498 Corynebac
831	36	34.6	250	7	ABO84368	AbO84368 Pseudomon	904	36	34.6	447	4	AAM38827	Aam38827 Human pol
832	36	34.6	260	6	ADX02026	Adx02026 SARS coro	905	36	34.6	447	5	AAE21720	Aae21720 Human PKI
833	36	34.6	266	8	ADK48784	Adk48784 Streptoco	906	36	34.6	447	7	ADM05915	Adm05915 Human pro
834	36	34.6	267	3	AAy81677	Aay81677 Streptoco	907	36	34.6	447	9	AEC88845	Aec88845 Human CDN
835	36	34.6	268	8	ABO58929	AbO58929 Human gen	908	36	34.6	451	3	AAG39206	Aag39206 Arabidops
836	36	34.6	270	5	AAU76160	Aau76160 Bacillus	909	36	34.6	453	7	ABO71768	AbO71768 Pseudomon
837	36	34.6	271	6	ABU00996	Abu00996 S. pneumo	910	36	34.6	453	8	ADN27283	Adn27283 Bacterial
838	36	34.6	280	2	AAE23178	Aae23178 Bacterial	911	36	34.6	454	8	ADS44648	Ads44648 Bacterial
839	36	34.6	280	2	AAE37292	Aae37292 Barley ty	912	36	34.6	456	6	ABP57494	Abp57494 Mycobacte
840	36	34.6	280	2	AAE63904	Aae63904 Type I ri	913	36	34.6	456	6	AAU16378	Aau16378 Murine pa
841	36	34.6	280	2	AAE74178	Aae74178 Barley Ty	914	36	34.6	457	6	AAU96704	Aau96704 Human nuc
842	36	34.6	280	2	AAW21698	Aaw21698 Barley tr	915	36	34.6	460	2	AAR69874	Aar69874 B.thermog
843	36	34.6	280	2	AAW25135	Aaw25135 Barley Tr	916	36	34.6	460	4	AAAM40613	Aam40613 Human pol
844	36	34.6	281	2	AAE25048	Aae25048 Protein s	917	36	34.6	460	6	AAO16374	Aao16374 Human pan
845	36	34.6	281	2	AAE52575	Aae52575 Protein s	918	36	34.6	466	3	AGC29656	Agc29656 Arabidops
846	36	34.6	292	6	ADA35355	Ada35355 Acinetoba	919	36	34.6	468	4	ABB61357	Abb61357 Drosophil
847	36	34.6	295	6	ABU65260	Abu65260 Novel hum	920	36	34.6	470	8	ADS27421	Ads27421 Bacterial
848	36	34.6	297	4	ABB65990	Abb65990 Drosophil	921	36	34.6	471	6	ABU45496	Abu45496 Protein e
849	36	34.6	300	8	ADX72798	Adx72798 Plant ful	922	36	34.6	472	3	AAU78103	Aau78103 Bacillus
850	36	34.6	302	10	AE97921	Aee97921 Mesorhizo	923	36	34.6	477	3	AGC42562	Agc42562 Arabidops
851	36	34.6	314	3	ABE53189	AbE53189 Macaca mu	924	36	34.6	479	5	ABP27212	Abp27212 Streptoco
852	36	34.6	324	3	ABB97327	Abb97327 Novel hum	925	36	34.6	479	8	ADV82130	Adv82130 Streptoco
853	36	34.6	320	3	AGB36746	Agb36746 Arabidops	926	36	34.6	483	2	AAW31538	Aaw31538 Amino aci
854	36	34.6	320	3	AAE19299	Aae19299 Arabidops	927	36	34.6	483	6	ABG74403	Abg74403 Banana 13
855	36	34.6	320	8	ADN72599	Adn72599 Thale cre	928	36	34.6	484	9	ABM92032	Abm92032 M. xanthu
856	36	34.6	321	3	AGA40095	AgA40095 Arabidops	929	36	34.6	492	8	ADN18407	Adn18407 Bacterial
857	36	34.6	324	8	ADX74837	Adx74837 Plant ful	930	36	34.6	493	6	ABU33590	Abu33590 Protein e
858	36	34.6	325	3	AAG40094	AgA40094 Arabidops	931	36	34.6	493	9	ABE41691	Aeb41691 L. pneumo
859	36	34.6	325	8	ADT60244	Adt60244 Plant pol	932	36	34.6	497	9	ABE38427	Aeb38427 L. pneumo
860	36	34.6	328	6	ABU30270	Abu30270 Protein e	933	36	34.6	504	8	ADN20868	Adn20868 Bacterial
861	36	34.6	336	6	ABU23886	Abu23886 Protein e	934	36	34.6	505	8	ADV88738	Adv88738 Streptoco
862	36	34.6	334	4	AAE82604	Aae82604 Cattle Ig	935	36	34.6	505	8	ADV79991	Adv79991 Streptoco
863	36	34.6	334	4	AAE82606	Aae82606 Sheep IgG	936	36	34.6	506	4	AGG78226	Agg78226 Lepomis c
864	36	34.6	337	2	ABB04127	Abb04127 Fusarium	937	36	34.6	507	3	ADJ29655	Adj29655 Arabidops
865	36	34.6	360	3	AGC42563	Agc42563 Arabidops	938	36	34.6	519	7	ADJ68604	Adj68604 Human hea
866	36	34.6	369	7	ABO81335	AbO81335 Pseudomon	939	36	34.6	519	8	ADQ37886	Adq37886 Variant h
867	36	34.6	371	6	ABU79127	Abu79127 Lipid A b	940	36	34.6	519	8	ADQ37881	Adq37881 Human DHP
868	36	34.6	371	7	ADF43387	Adf43387 Lipid A b	941	36	34.6	531	4	ABB63988	Abb63988 Drosophil
869	36	34.6	371	8	ADT49844	Adt49844 Murine TA	942	36	34.6	539	3	AAG29654	Aag29654 Arabidops
870	36	34.6	371	9	AEA03058	Aea03058 Lipid A b	943	36	34.6	540	2	AAE72678	Aae72678 Protein d
871	36	34.6	373	2	AAW88213	Aaw88213 Pseudomon	944	36	34.6	549	4	ABB71680	Abb71680 Drosophil
872	36	34.6	374	7	ADC01328	Adc01328 Enterobae	945	36	34.6	556	7	ABE80720	Abe80720 Microgate
873	36	34.6	374	8	ADH88970	Adh88970 Escherich	946	36	34.6	557	2	AAW57844	Aaw57844 F24A muta
874	36	34.6	376	2	AAE15272	Aae15272 Fusarium	947	36	34.6	557	2	AAW57840	Aaw57840 Wild type
875	36	34.6	376	2	AAE25429	Aae25429 Cellulase	948	36	34.6	557	2	AAW57846	Aaw57846 F24L muta
876	36	34.6	376	2	AAE25527	Aae25527 Fusarium	949	36	34.6	557	2	AAW57845	Aaw57845 F24V muta
877	36	34.6	376	2	AAE25466	Aae25466 Endogluca	950	36	34.6	557	2	AAW57843	Aaw57843 V56T muta
878	36	34.6	376	2	AAE27969	Aae27969 Endogluca	951	36	34.6	557	7	ADA38225	Ada38225 Escherich
879	36	34.6	376	2	AAE24064	Aae24064 Endogluca	952	36	34.6	566	8	ADS23484	Ads23484 Bacterial
880	36	34.6	376	2	AAE37151	Aae37151 Dye trans	953	36	34.6	570	4	ABB63990	Abb63990 Drosophil
881	36	34.6	376	2	AAE67389	Aae67389 F. oxyspo	954	36	34.6	571	4	ABG00360	Abg00360 Novel hum
882	36	34.6	376	2	AAW46617	Aaw46617 Fusarium	955	36	34.6	582	4	AAE78227	Aae78227 Lepomis c
883	36	34.6	378	8	ADT49883	Adt49883 Murine TA	956	36	34.6	588	9	ADX06294	Adx06294 Cyclin-de
884	36	34.6	401	7	ABO84231	AbO84231 Pseudomon	957	36	34.6	599	4	ABB71947	Abb71947 Drosophil
885	36	34.6	406	9	AEC95648	Aec95648 Calcium c	958	36	34.6	599	9	AEA00258	Aea00258 Leishmani
886	36	34.6	409	2	AAW01501	Aaw01501 50 kD end	959	36	34.6	603	5	ADL23399	Adl23399 Plant AMP
887	36	34.6	409	2	AAW17927	Aaw17927 Abrading	960	36	34.6	603	7	ADL72426	Adl72426 Arabidops
888	36	34.6	411	2	AAE83399	Aae83399 Fusarium	961	36	34.6	603	7	ADL72277	Adl72277 Arabidops
889	36	34.6	420	8	ADS21028	Ads21028 Bacterial	962	36	34.6	605	7	ABM04806	Abm04806 Rat ribop
890	36	34.6	423	6	ABP80988	Abp80988 N. gonorr	963	36	34.6	605	7	ADE63298	Ade63298 Rat Prote
891	36	34.6	424	6	ABU24236	Abu24236 Protein e	964	36	34.6	605	8	ADT57496	Adt57496 Plant pol
892	36	34.6	425	4	ABG29201	Abg29201 Novel hum	965	36	34.6	607	8	ADM57739	Adm57739 Human cal
893	36	34.6	425	8	ADS30686	Ads30686 Bacterial	966	36	34.6	608	9	ADW25843	Adw25843 Hybrid al
894	36	34.6	426	6	ABU19824	Abu19824 Protein'e	967	36	34.6	608	9	ADW25847	Adw25847 Hybrid al
895	36	34.6	427	2	AAU15239	Aau15239 Fusarium	968	36	34.6	608	9	ADW25846	Adw25846 Hybrid al
896	36	34.6	433	2	AAW12400	Aaw12400 A. chryso	969	36	34.6	608	9	ADW25851	Adw25851 Hybrid al
897	36	34.6	437	4	AAW79064	Aaw79064 Corynebac	970	36	34.6	608	9	AAW50111	Aaw50111 Feline pa
898	36	34.6	437	4	AAU71863	Aau71863 C. glutam	971	36	34.6	620	4	AAU37023	Aau37023 Staphyloc
899	36	34.6	437	4	AAE90472	Aae90472 C glutami	972	36	34.6	628	4	AAU37023	Aau37023 Staphyloc

973 36 34.6 628 6 ABU16319 Protein e
 974 36 34.6 628 6 ABM73039 Staphyloc
 975 36 34.6 631 8 ADY07077 Plant ful
 976 36 34.6 634 2 AAR23069 GL-7ACA a
 977 36 34.6 640 4 AAU33800 Staphyloc
 978 36 34.6 641 8 ADX88667 Plant ful
 979 36 34.6 655 4 AAG90809 C glutami
 980 36 34.6 656 7 ABM87997 Rice abio
 981 36 34.6 656 8 ADP83535 Breast sp
 982 36 34.6 657 7 ADD13583 C. glutam
 983 36 34.6 666 8 ADN17365 Bacterial
 984 36 34.6 686 3 AAG42561 Arabidops
 985 36 34.6 693 8 ADX76161 Plant ful
 986 36 34.6 694 2 ABY19572 Protein e
 987 36 34.6 708 2 AAU34501 Porphorym
 988 36 34.6 712 8 ADQ19271 Human sof
 989 36 34.6 724 3 AAB56475 Human pro
 990 36 34.6 762 4 ABB64096 Drosophil
 991 36 34.6 772 2 AAY34374 Porphorym
 992 36 34.6 774 7 ADF04771 Bacterial
 993 36 34.6 814 7 ADJ71142 Human hea
 994 36 34.6 820 8 ABM82800 Human dia
 995 36 34.6 821 7 ADM26833 Hyperther
 996 36 34.6 841 6 ABM70198 Photorhab
 997 36 34.6 846 1 AAP61009 Sequence
 998 36 34.6 846 4 AAB37817 A. faecal
 999 36 34.6 846 7 ADA38223 Escherich
 1000 36 34.6 863 4 AAB95203 Human pro

ALIGNMENTS

RESULT 1
 ADR88213
 ID ADR88213 standard; peptide; 19 AA.
 AC ADR88213;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human heparanase epitope p8 #7.
 XX
 KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
 KW human; heparanase; enzyme; epitope.
 XX
 OS Homo sapiens.
 XX
 PN US20041706311-A1.
 XX
 PD 02-SEP-2004.
 XX
 PF 28-NOV-2003; 2003US-00722502.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 PR 22-AUG-2003; 2003US-00645659.
 XX
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRG/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (VGEL/) VAN GELDER J M.
 PA (VLOD/) VLODAVSKY I.

(FRIE/) FRIEDMANN Y.
 PA Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 PI Friedmann Y;
 XX
 DR WPI; 2004-625084/60.
 XX
 PT Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.
 PT
 CC Claim 7; SEQ ID NO 7; 58pp; English.
 XX
 CC The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is human heparanase epitope.
 XX
 SQ Sequence 19 AA;
 Query Match 100.0%; Score 104; DB 8; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 DB 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 RESULT 2
 ADT78180
 ID ADT78180 standard; peptide; 19 AA.
 XX
 AC ADT78180;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE Functional peptide epitope of human heparanase, p8#7.
 XX
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.
 OS Homo sapiens.

XX AEA42429 standard; peptide; 19 AA.
 PN US2004213789-A1.
 XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.
 PF 02-SEP-1997; 97US-00922170.
 XX 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.

XX (YACO) YACOB-ZEEVI O.
 PA (PERE) PERETZ T.
 PA (MIRO) MIRON D.
 PA (SHLO) SHLOMI Y.
 PA (PECK) PECKER I.
 PA (AYAL) AYAL-HERSHKOVITZ M.
 PA (FEIN) FEINSTEIN E.
 PA (GELD) GELDER J M V.
 PA (VLOD) VLODAVSKY I.
 PA (FRIE) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;
 XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX Claim 53; SEQ ID NO 7; 68pp; English.
 PS The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents a functional peptide
 CC epitope of human heparanase.

XX Sequence 19 AA;

Query Match 100.0%; Score 104; DB 8; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19

Db 1 PAYLRFGGTKTDFLIFDPK 19

RESULT 3

AEA42429
 ID AEA42429 standard; peptide; 19 AA.

XX AEA42429;

XX 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:7.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotrophic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

XX Homo sapiens.

XX AU2004201462-A1.

XX 06-MAY-2004.

XX 08-APR-2004; 2004AU-00201462.

XX 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.

XX Claim 7; SEQ ID NO 7; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents a human
 CC heparanase epitope peptide, which is used in the exemplification of the
 CC present invention.

XX Sequence 19 AA;

Query Match 100.0%; Score 104; DB 9; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19

|||||
1 PAYLRFGGTKDFLIFDPK 19

Db
RESULT 4
ADR88217
ID ADR88217 standard; protein; 74 AA.
AC
XX
ACR88217;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human mature processed heparanase dimer 8 kDa subunit.
XX
KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
KW human; heparanase; enzyme.
XX
OS Homo sapiens.
XX
XX US2004170631-A1.
PN
XX
XX 02-SEP-2004.
PD
XX
PF 28-NOV-2003; 2003US-00722502.
XX
XX 02-SEP-1997; 97US-00922170.
PR
XX 01-MAY-1998; 98US-00071739.
PR
XX 04-NOV-1998; 98US-00186200.
PR
XX 19-FEB-2003; 2003US-00368044.
PR
XX 22-AUG-2003; 2003US-00645659.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VGLD/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI; 2004-625084/60.
DR
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX
XX Claim 2; SEQ ID NO 11; 58pp; English.
XX
XX The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an antibody or its portion
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease), for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a

CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is the 8 kDa subunit of human mature
XX processed heparanase dimer.
SQ Sequence 74 AA;
Query Match 100.0%; Score 104; DB 8; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.4e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PAYLRFGGTKDFLIFDPK 19
Db 54 PAYLRFGGTKDFLIFDPK 72
RESULT 5
ADR78184
ID ADT78184 standard; protein; 74 AA.
XX
XX AC ADT78184;
XX
XX DT 13-JAN-2005 (first entry)
XX
XX DE 8kDa subunit of mature processed human heparanase dimer.
XX
XX KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cystic; antiinflammatory; vulnary; antiarteriosclerotic;
KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.
XX
XX OS Homo sapiens.
XX
XX PN US2004213789-A1.
XX
XX PD 28-OCT-2004.
XX
XX PF 22-AUG-2003; 2003US-00645659.
XX
XX PR 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (GELD/) GELDER J M V.
PA (VLOD/) VLODAVSKY I.
XX (FRIE/) FRIEDMANN Y.
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX WPI; 2004-774790/76.
DR
XX
XX New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
PS Claim 5; SEQ ID NO 11; 68pp; English.
XX
XX The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents the 8kDa subunit of
CC mature processed human heparanase dimer.
XX
SQ Sequence 74 AA;

Query Match 100.0%; Score 104; DB 8; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.4e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19
Db |||||
54 PAYLRFGGTKTDFLIFDPK 72

RESULT 6
ADY27059
ID ADY27059 standard; protein; 74 AA.
XX
XX AC ADY27059;
XX
DT 05-MAY-2005 (first entry)
XX
DE Heparanase inhibitor protein #2.
XX
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.
XX
OS Homo sapiens.
XX
XX WO2005016227-A2.
PN
XX 24-FEB-2005.
PD
XX 12-AUG-2004; 2004WO-IL000744.
PF
XX 14-AUG-2003; 2003US-0494800P.
PR
XX 12-JAN-2004; 2004US-0535492P.
XX

PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX
PI Van-Gelder JM, Miron D;
XX
DR WPI; 2005-182203/19.
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.
XX
PS Claim 55; SEQ ID NO 35; 211pp; English.
XX
XX The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the
CC invention.
XX
SQ Sequence 74 AA;

Query Match 100.0%; Score 104; DB 9; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.4e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19
Db |||||
54 PAYLRFGGTKTDFLIFDPK 72

RESULT 7
ADZ18994
ID ADZ18994 standard; protein; 74 AA.
XX
XX AC ADZ18994;
XX
DT 16-JUN-2005 (first entry)
XX
DE Human heparanase consensus cleavage site #1.
XX
KW Enzyme engineering; heparanase; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Homo sapiens.
XX
XX WO2005030962-A1.
PN
XX 07-APR-2005.
PD
XX 17-SEP-2004; 2004WO-EP010517.
PF
XX 26-SEP-2003; 2003US-0506479P.
PR
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHÉ BIOL MOLECOLARE ANGELETTI.
PA
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
PI
XX

DR WPI; 2005-273382/28.
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Disclosure; SEQ ID NO 15; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a human heparanase
CC consensus cleavage site used in the scope of the invention.
XX
XX Sequence 74 AA;
SQ
Query Match 100.0%; Score 104; DB 9; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.4e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDLIFDPK 19
Db ||||||||||||||||
54 PAYLRFGGTKTDLIFDPK 72
RESULT 8
ID AEA42433
AC AEA42433 standard; protein; 74 AA.
XX
XX AEA42433;
DT 28-JUL-2005 (first entry)
XX
DE Human mature heparanase dimer 8 kDa subunit SEQ ID NO:11.
XX
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
KW angiogenesis disorder; cancer; tumor; metastasis.
XX
XX Homo sapiens.
OS
XX AU2004201462-A1.
PN
XX 06-MAY-2004.
PD
XX 08-APR-2004; 2004AU-00201462.
PF
XX 08-APR-2004; 2004AU-00201462.
PR
XX (INST-) INSIGHT BIOPHARMACEUTICALS LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinstein E;
XX
XX WPI; 2005-173343/19.
DR
XX Novel isolated antibody capable of specifically binding to epitope of
XX heparanase protein, useful for preventing and treating heparanase-related
PT disorder such as inflammatory disorder, scars, autoimmune conditions or
PT angiogenesis.
PT
XX Claim 2; SEQ ID NO 11; 260pp; English.
PS
XX The invention relates to an isolated antibody or its portion (I) capable
XX of specifically binding to an epitope of a heparanase protein. Also
CC described: (1) a cell line (II) for producing a monoclonal antibody or

CC its portion, comprising a cell line for producing (I); (2) a
CC pharmaceutical composition comprising (I) and a carrier; and (3) an
CC affinity medium (III) for binding human heparanase polypeptides,
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
CC useful for treating a subject suffering from a pathological condition,
CC which involves administering (I) to the subject. (I) is useful for
CC preventing and treating heparanase-related disorder or condition chosen
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
CC cell proliferation, invasion of circulating tumor cells and metastatic
CC disease. (II) is useful for detecting the presence of heparanase
CC polypeptide in a sample. (I) is useful for detecting heparanase-related
CC disease or condition in a subject such as vertebrate, preferably mammal
CC e.g., human. The heparanase-related disorder or condition further
CC includes renal disease or disorder chosen from diabetic nephropathy,
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
CC and renal cell carcinoma. The present sequence represents the 8 kDa
CC subunit of the human mature processed heparanase dimer, which is used in
CC the exemplification of the present invention.
XX
XX Sequence 74 AA;
SQ
Query Match 100.0%; Score 104; DB 9; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.4e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDLIFDPK 19
Db ||||||||||||||||
54 PAYLRFGGTKTDLIFDPK 72
RESULT 9
ID ADZ19012 standard; protein; 174 AA.
ADZ19012
XX
XX AC ADZ19012;
XX
XX 16-JUN-2005 (first entry)
DT
XX
DE Human heparanase protein.
XX
XX Enzyme engineering; heparanase; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
XX Homo sapiens.
OS
XX WO2005030962-A1.
PN
XX 07-APR-2005.
PD
XX 17-SEP-2004; 2004WO-EP010517.
PF
XX 26-SEP-2003; 2003US-0506479P.
PR
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
PA
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
PI
XX WPI; 2005-273382/28.
DR
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Disclosure; SEQ ID NO 33; 65pp; English.
PS
XX The invention relates to a synthetic nucleic acid molecule that encodes
XX mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a human heparanase
CC consensus cleavage site used in the scope of the invention.
XX
XX Sequence 74 AA;
SQ

CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a human heparanase protein
CC used in the scope of the invention.

XX SQ Sequence 174 AA;
XX Query Match 100.0%; Score 104; DB 9; Length 174;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLFDPK 19
| | | | | | | | | | | | | | | | | | | | | |
Db 55 PAYLRFGGTKTDFLFDPK 73

RESULT 10
ID ADY27061
XX ADY27061 standard; protein; 460 AA.
AC ADY27061;
XX 05-MAY-2005 (first entry)
DT Heparanase inhibitor protein #4.
DE Heparanase inhibitor protein #4.
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.

XX Homo sapiens.
OS
XX WO2005016227-A2.
XX 24-FEB-2005.
XX 12-AUG-2004; 2004WO-IL0000744.
XX 14-AUG-2003; 2003US-0494800P.
PR 12-JAN-2004; 2004US-0535492P.
XX (NSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA Van-Gelder JM, Miron D;
PI WPI; 2005-182203/19.
DR
XX Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.

XX Disclosure; SEQ ID NO 37; 211pp; English.
XX The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the

CC invention.
XX SQ Sequence 460 AA;
XX Query Match 100.0%; Score 104; DB 9; Length 460;
XX Best Local Similarity 100.0%; Pred. No. 5.8e-09;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLFDPK 19
| | | | | | | | | | | | | | | | | | | | | |
Db 54 PAYLRFGGTKTDFLFDPK 72

RESULT 11
ID AEB87589
XX AEB87589 standard; protein; 486 AA.
AC AEB87589;
XX 06-OCT-2005 (first entry)
DT Human heparanase 65delta20 deletion mutant.

XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; mutein.

XX Homo sapiens.
OS Synthetic.
XX WO2005071070-A2.
XX 04-AUG-2005.
XX 20-JAN-2005; 2005WO-IL0000069.
XX 22-JAN-2004; 2004IL-00160025.
PR 28-JUL-2004; 2004US-00901943.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

PA Vlodavsky I, Ilan N, Levy-Adam F;
PI WPI; 2005-564219/57.
DR N-PSDB; AEB87588.
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX Claim 107; SEQ ID NO 31; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid

CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX
 SQ Sequence 486 AA;

Query Match 100.0%; Score 104; DB 9; Length 486;
 Best Local Similarity 100.0%; Pred. No. 6.1e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKDFLIFDPK 19

|||||

54 PAYLRFGGTKDFLIFDPK 72

RESULT 12

ADZ18996
 ID ADZ18996 standard; protein; 492 AA.

XX
 AC ADZ18996;

DT 16-JUN-2005 (first entry)

XX Hep106 construct protein.

DE
 XX Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.

XX Synthetic.

OS
 XX WO2005030962-A1.

PN 07-APR-2005.

PD 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

PR 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

PA Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

PI WPI; 2005-273382/28.

XX N-PSDB; ADZ18997.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.

XX Example 2; SEQ ID NO 17; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hep106 construct protein
 CC used in the scope of the invention.

XX Sequence 492 AA;

Query Match 100.0%; Score 104; DB 9; Length 492;
 Best Local Similarity 100.0%; Pred. No. 6.2e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKDFLIFDPK 19

|||||

89 PAYLRFGGTKDFLIFDPK 107

RESULT 13

AE8B7562
 ID AEB87562 standard; protein; 493 AA.
 XX
 AC AEB87562;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE Human heparanase 65delta15 deletion mutant.

XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.

XX Homo sapiens.

OS Synthetic.

XX WO2005071070-A2.

XX 04-AUG-2005.

XX 20-JAN-2005; 2005WO-IL000068.

XX 22-JAN-2004; 2004IL-00160025.

PR 28-JUL-2004; 2004US-00901943.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

PA Vlodavsky I, Ilan N, Levy-Adam F;

XX WPI; 2005-564219/57.

DR N-PSDB; AEB87561.

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.

XX Claim 105; SEQ ID NO 4; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.

XX Sequence 493 AA;

Query Match 100.0%; Score 104; DB 9; Length 493;
 Best Local Similarity 100.0%; Pred. No. 6.2e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKDFLIFDPK 19

|||||

54 PAYLRFGGTKDFLIFDPK 72

RESULT 14

ADZ18999
ID ADZ18999 standard; protein; 495 AA.
XX
AC ADZ18999;
XX
DT 16-JUN-2005 (first entry)
XX
DE Hep109 construct protein.
XX
KW Enzyme engineering; heparanase; hep109; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
PA (RICE-) 1ST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
DR WPI; 2005-273382/28.
DR N-PSDB; ADZ18998.
XX
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
PS Example 2; SEQ ID NO 20; 65pp; English.
XX
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hep109 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 495 AA;
Query Match 100.0%; Score 104; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDPK 19
DB 89 PAYLRFGGTKTDFLIFDPK 107
RESULT 15
AEB87587
ID AEB87587 standard; protein; 497 AA.
XX
AC AEB87587;
XX
DT 06-OCT-2005 (first entry)
XX
DE Human heparanase 65delta10 deletion mutant.
XX
KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
XX

KW autoimmune disease; immunosuppressive; drug screening; mutein.
XX Homo sapiens.
OS Synthetic.
XX
PN WO2005071070-A2.
XX
PD 04-AUG-2005.
XX
PF 20-JAN-2005; 2005WO-IL000068.
XX
PR 22-JAN-2004; 2004IL-00160025.
PR 28-JUL-2004; 2004US-00901943.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
PI Vlodavsky I, Ilan N, Levy-Adam F;
XX
DR WPI; 2005-564219/57.
DR N-PSDB; AEB87586.
XX
PT New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX
PS Claim 106; SEQ ID NO 29; 167pp; English.
XX
CC The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
SQ Sequence 497 AA;
Query Match 100.0%; Score 104; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDPK 19
DB 54 PAYLRFGGTKTDFLIFDPK 72
RESULT 16
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX
AC ADZ19000;
XX
DT 16-JUN-2005 (first entry)
XX
DE HepGS3 construct protein.
XX
KW Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.

XX WO2005030962-A1.
 XX 07-APR-2005.
 XX 17-SEP-2004; 2004WO-EP010517.
 XX 26-SEP-2003; 2003US-0506479P.
 XX 20-JAN-2004; 2004US-0537729P.
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
 XX WPI; 2005-273382/28.
 XX N-PSDB; ADZ19001.
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX Example 2; SEQ ID NO 21; 65pp; English.
 XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoprotease. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepgS3 construct protein
 CC used in the scope of the invention.
 XX Sequence 501 AA;
 SQ

Query Match 100.0%; Score 104; DB 9; Length 501;
 Best Local Similarity 100.0%; Pred. No. 6.3e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 17
 ADZ19005
 ID ADZ19005 standard; protein; 507 AA.
 XX AC
 XX ADZ19005;
 XX 16-JUN-2005 (first entry)
 XX HepG6 construct protein.
 XX Enzyme engineering; heparanase; hepgS6; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX Synthetic.
 XX OS
 XX WO2005030962-A1.
 XX 07-APR-2005.
 XX 17-SEP-2004; 2004WO-EP010517.
 XX 26-SEP-2003; 2003US-0506479P.
 XX 20-JAN-2004; 2004US-0537729P.
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
 XX

DR WPI; 2005-273382/28.
 XX N-ESDB; ADZ19003.
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX Example 2; SEQ ID NO 26; 65pp; English.
 XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoprotease. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepgS6 construct protein
 CC used in the scope of the invention.
 XX Sequence 507 AA;
 SQ

Query Match 100.0%; Score 104; DB 9; Length 507;
 Best Local Similarity 100.0%; Pred. No. 6.4e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 18
 ADY27058
 ID ADY27058 standard; protein; 508 AA.
 XX AC
 XX ADY27058;
 XX 05-MAY-2005 (first entry)
 XX Human inactive heparanase protein.
 XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.
 XX Homo sapiens.
 XX OS
 XX WO2005016227-A2.
 XX 24-FEB-2005.
 XX 12-AUG-2004; 2004WO-IL000744.
 XX 14-AUG-2003; 2003US-0494800P.
 XX 12-JAN-2004; 2004US-0535492P.
 XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
 XX Van Gelder JM, Miron D;
 XX WPI; 2005-182203/19.
 XX Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.
 XX Claim 257; SEQ ID NO 34; 211pp; English.
 XX The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin

Query Match 100.0%; Score 104; DB 5; Length 527;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDLIFDPK 19
|||||
Db 73 PAYLRFGGTKTDLIFDPK 91

RESULT 21
ABW02018
ID ABW02018 standard; protein; 527 AA.
XX
AC ABW02018;
XX
DT 12-FEB-2004 (first entry)
XX
DE Chimeric human-chicken heparanase protein.
XX
KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
KW immunosuppressive; vulnary; bactericide; anti-angiogenic; virucide;
KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
KW enzyme; human.
XX
OS Chimeric - Gallus gallus.
OS Chimeric - Homo sapiens.
XX
XX US2003180788-A1.
XX
PD 25-SEP-2003.
XX
XX 08-MAY-2003; 2003US-00431438.
XX
PR 20-SEP-2000; 2000US-00666390.
PR 16-AUG-2001; 2001US-00930218.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX WPI: 2003-843931/78.
XX N-PSDB; AAD63532.
XX
PT Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
PT for treating cancers, microbial infections and aiding wound healing.
XX
XX Example; Page 26-28; Opp; English.
XX
CC The present invention relates to novel jungle red fowl heparanase protein
CC and polynucleotides encoding such proteins. Heparanase sequences can be
CC used to develop treatments for various diseases, to develop diagnostic
CC assays for these diseases and to provide new tools for basic and directed
CC research especially in the fields of medicine and biology. They can be
CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
CC and autoimmunity. Recombinant heparanase offers a potential treatment for
CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
CC some bacterial and protozoa infections. Recombinant heparanase can also
CC be used to neutralise plasma heparin, as a potential replacement of
CC protamine. Sequences of the invention are used in protein therapy. The
CC present sequence is chimeric human-chicken heparanase protein
XX
SQ Sequence 527 AA;

Query Match 100.0%; Score 104; DB 7; Length 527;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDLIFDPK 19
|||||
Db 73 PAYLRFGGTKTDLIFDPK 91

RESULT 22
ADO63825
ID ADO63825 standard; protein; 527 AA.
XX
AC ADO63825;
XX
DT 26-AUG-2004 (first entry)
XX
DE Chimeric heparanase mutant E225A, SEQ ID:10.
XX
KW Human; chicken; heparanase; heparanase-derived protein;
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; IAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnary; mutant; mutein.
XX
OS Homo sapiens.
OS Gallus gallus.
OS Synthetic.
OS Chimeric.
XX
XX Key Location/Qualifiers
XX Peptide 1..18
XX Region 19..527
XX Misc-difference 209
XX /note= "Chicken heparanase signal peptide"
XX /note= "Corresponds to residues 35-543 of human
XX heparanase mutant E225A (SEQ ID NO:7)"
XX /note= "Ala replaces wild-type Glu (active site proton
XX donor). Corresponds to residue 225 of human heparanase
XX mutant E225A (SEQ ID NO:7)"
XX Active-site 327
XX /note= "Active site nucleophile"
XX
XX WO2004048558-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-IL000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;
XX
XX WPI: 2004-450373/42.
XX N-PSDB; ADO63819.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Claim 10; SEQ ID NO 10; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The
XX invention also relates to the heparanase-derived proteins and host cells
XX comprising the nucleic acid constructs of the invention. The heparanase-
XX derived proteins are especially mutants of human heparanase in which the
XX active site proton donor Glu225 and/or the active site nucleophile Glu343
XX are replaced with Ala (ADO63822-ADO63824), and the proteins may
XX optionally further comprise an avian heparanase signal peptide (ADO63825-


```

PF 31-AUG-1998; 98WO-US017954.
XX
XX 02-SEP-1997; 97US-00922170.
XX 02-JUL-1998; 98US-00109386.
XX
XX (INST-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI; 1999-302255/25.
DR N-PSDB; AAX35648.
XX
XX New human polynucleotide useful for treating angiogenesis, restenosis,
PT and inflammation.
XX
XX Claim 6; Fig 1; 63pp; English.
XX
XX The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of heparin-
CC binding growth factors, cellular responses to heparin-binding growth
CC factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections or
CC disintegration of neurodegenerative plaques. Heparanase may be useful for
CC conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and
CC renal failure in biopsy specimens, plasma samples, and body fluids. The
CC present sequence represents human heparanase
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 104; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTTKDFLIFDPK 19
Db ||||| ||||| ||||| ||||| |||||
89 PAYLRFGGTTKDFLIFDPK 107

RESULT 29
AAY17082
ID AAY17082 standard; protein; 543 AA.
XX
XX AAY17082;
XX
XX 21-JUL-1999 (first entry)
XX
XX Human heparanase enzyme.
XX
XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.
XX
XX Homo sapiens.
XX
XX WO9921975-A1.
XX
XX 06-MAY-1999.
XX
XX 28-OCT-1998; 98WO-AU000898.
XX
XX 28-OCT-1997; 97AU-00000062.
XX 09-DEC-1997; 97AU-00000812.
XX
XX (AUSU ) UNIV AUSTRALIAN NAT.
XX
XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

PF 31-AUG-1998; 98WO-US017954.
XX
XX 02-SEP-1997; 97US-00922170.
XX 02-JUL-1998; 98US-00109386.
XX
XX (INST-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI; 1999-302255/25.
DR N-PSDB; AAX35648.
XX
XX New human polynucleotide useful for treating angiogenesis, restenosis,
PT and inflammation.
XX
XX Claim 6; Fig 1; 63pp; English.
XX
XX The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a human heparanase
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 104; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTTKDFLIFDPK 19
Db ||||| ||||| ||||| ||||| |||||
89 PAYLRFGGTTKDFLIFDPK 107

RESULT 30
AAY57590
ID AAY57590 standard; protein; 543 AA.
XX
XX AAY57590;
XX
XX 02-MAR-2000 (first entry)
XX
XX Human heparanase.
XX
XX Human; heparanase; hpa; genetic modification; expression; anticancer;
KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
KW micrometastasis; autoimmune lesion; kidney failure.
XX
XX Homo sapiens.
XX
XX WO9957244-A1.
XX
XX 11-NOV-1999.
XX
XX 29-APR-1999; 99WO-US009256.
XX
XX 01-MAY-1998; 98US-00071618.
XX 02-MAR-1999; 99US-00260038.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (FRIE/) FRIEDMAN M M.
XX
XX Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;
XX Peleg Y, Shlomi Y;
XX
XX WPI; 2000-062144/05.
XX N-PSDB; AAZ39195.
XX

```


CC heparanase, for diagnosis and monitoring of diseases (especially
 CC metastasis), for treatment of heparanase-associated diseases (e.g.
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
 CC sepsis and inflammatory or autoimmune disease), for targeted drug
 CC delivery (e.g. of anticancer agents) and as research reagents. The
 CC present sequence represents human heparanase, which is used in the
 CC exemplification of the present invention
 XX
 XX Sequence 543 AA;

Query Match 100.0%; Score 104; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 6.9e-09; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 33
 AAY97635
 ID AAY97635 standard; protein; 543 AA.

XX
 AC AAY97635;

XX
 DT 20-APR-2001 (first entry)

XX
 DE Human heparanase protein sequence.

XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
 KW gene therapy; human.

XX Homo sapiens.

OS WO200100643-A2.

PN 04-JAN-2001.

XX
 PF 19-JUN-2000; 2000WO-IL000358.

XX
 PR 25-JUN-1999; 99US-0140801P.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX
 PI Pecker I, Michal I, Itzhaki H;

XX
 DR WPI; 2001-137930/14.

XX New polynucleotides and polypeptides that are distantly homologous to
 PT heparanase, useful in wound healing, as well as in gene therapy protocols
 PT for angiogenesis, restenosis, atherosclerosis, or inflammation.

XX
 PS Disclosure; Page 64-65; 67pp; English.

XX This sequence represents a heparanase of the invention. The heparanase
 CC DNA and protein sequences are useful in wound healing, angiogenesis,
 CC restenosis, atherosclerosis, inflammation, pulmonary diseases,
 CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and
 CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
 CC sequence is particularly useful in gene therapy

XX
 SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 4; Length 543;
 Best Local Similarity 100.0%; Pred. No. 6.9e-09; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 34

AAB86206
 ID AAB86206 standard; protein; 543 AA.

XX
 AC AAB86206;

XX
 DT 24-AUG-2001 (first entry)

XX
 DE Human heparanase inhibitor protein.

XX Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
 KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
 KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
 KW hepatic congestion; hydrothorax; ascite; nocturia; human.

OS Homo sapiens.

XX
 PN DE19955803-A1.

XX
 PD 23-MAY-2001.

XX
 PF 19-NOV-1999; 99DE-01055803.

XX
 PR 19-NOV-1999; 99DE-01055803.

XX
 PA (KNOL) KNOLL AG.

XX
 PI Herr D, Hahn A, Laux V;

XX
 DR WPI; 2001-368371/39.

XX
 DR N-PSDB; AAH20940.

XX Treatment or prevention of cardiac insufficiency and related conditions,
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of
 PT heparanase inhibitor.

XX
 PS Disclosure; Page 11-13; 16pp; German.

XX This invention describes a novel heparanase inhibitor which can be used
 CC for the treatment or prevention of cardiac insufficiency and associated
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
 CC the invention has cardiant, nephrotropic and hepatotropic activity. The
 CC products of the invention can be used in human and veterinary medicine,
 CC for the treatment or prevention of congestive heart failure e.g. primary
 CC cardiomyopathy. Associated conditions treated or prevented with the
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
 CC nocturia can also be treated. This sequence represents the human
 CC heparanase protein described in the method of the invention

XX
 SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 4; Length 543;
 Best Local Similarity 100.0%; Pred. No. 6.9e-09; Indels 0; Gaps 0;
 Matches 19; Conservative 0; Mismatches 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 35

AAB88361
 ID AAB88361 standard; protein; 543 AA.

XX
 AC AAB88361;

XX
 DT 23-MAY-2001 (first entry)

XX DE Human membrane or secretory protein clone PSEC0090.
XX DE
XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;
XX KW rheumatoid arthritis; diabetes.
XX OS Homo sapiens.
XX PN EP1067182-A2.
XX PD 10-JAN-2001.
XX PF 07-JUL-2000; 2000EP-00114090.
XX PR 08-JUL-1999; 99JP-00194179.
XX PR 11-JAN-2000; 2000JP-00118775.
XX PR 02-MAY-2000; 2000JP-001183766.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX WPI; 2001-093989/11.
XX DR N-PSDB; AAF93788.
XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX PT gene therapy or as candidate target molecules in drug development.
XX PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.
XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
XX CC which encode human secretory or membrane proteins represented by AAF93917
XX CC - AAF88419. Included in the invention are primers AAF93917 - AAF94295 and
XX CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
XX CC invention. The invention also includes methods for the production of
XX CC antibodies directed against the proteins, and cDNA sequences, which can
XX CC be used in vaccines. The polynucleotide sequences can be used in gene
XX CC therapy. The polynucleotide sequences and the proteins they encode may be
XX CC used in the prevention, treatment and diagnosis of diseases associated
XX CC with inappropriate secretory protein/membrane protein expression. The
XX CC nucleic acids and complementary sequences may also be used as DNA probes
XX CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
XX CC and quantitate the presence of similar nucleic acid sequences in samples.
XX CC They may also be used to study the expression and function of secretory
XX CC proteins/membrane polypeptides and their role in metabolism. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC against them and in assays to identify modulators (agonists and
XX CC antagonists) of expression and activity. The antibodies and antagonists
XX CC may also be used as therapeutic agents to down regulate expression and
XX CC activity. The antibodies may also be used as diagnostic agents for
XX CC detecting the presence of the polypeptides in samples (e.g. by enzyme
XX CC linked immunosorbant assay (ELISA)). Examples of diseases which may be
XX CC treated include rheumatoid arthritis and diabetes
XX SQ
XX Sequence 543 AA;
Query Match 100.0%; Score 104; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDPK 19
Db 89 PAYLRFGGTKTDFLIFDPK 107
RESULT 36
ABB07813
ID ABB07813 standard; protein; 543 AA.
XX AC ABB07813;
XX AC
XX DT 03-JUL-2002 (first entry)
XX

DE Human heparanase sequence.
XX KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
XX KW anti-protozoan; neuroprotective; heparin; human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..35
XX FT /note= "signal peptide"
XX FT Protein 36..543
XX FT /note= "mature protein"
XX PN US2002034810-A1.
XX XX
XX PD 21-MAR-2002.
XX PF 16-AUG-2001; 2001US-00930218.
XX PR 20-SEP-2000; 2000US-00666390.
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PI Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;
XX WPI; 2002-338926/37.
XX PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
XX PT to treat various heparin-related disorders and the signal peptide is
XX PT useful in production of membrane-targeted or secreted recombinant
XX PT proteins.
XX PS Disclosure; Fig 1a; 39pp; English.
XX CC The invention relates to an isolated avian and reptile nucleic acid,
XX CC encoding a polypeptide with heparanase catalytic activity. The signal
XX CC peptide of the nucleic acid can be used to express membrane-associated or
XX CC secreted proteins in heterologous expression systems. The encoded
XX CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
XX CC invasion, and to intervene with pathologies associated with impaired
XX CC heparin-binding growth factors, cellular responses to heparin-binding
XX CC growth factors and cytokines, cell interaction with plasma lipoproteins,
XX CC cellular susceptibility to viral, protozoa and bacterial infections or
XX CC disintegration of neurodegenerative plaques. The present sequence
XX CC represents a human heparanase protein sequence used in similarity studies
XX SQ
XX Sequence 543 AA;
Query Match 100.0%; Score 104; DB 5; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDPK 19
Db 89 PAYLRFGGTKTDFLIFDPK 107
RESULT 37
ADD18950
ID ADD18950 standard; protein; 543 AA.
XX AC ADD18950;
XX AC
XX DT 15-JAN-2004 (first entry)
XX DE Human disease related protein SeqID439.
XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
XX KW antiarteriosclerotic; vulnery; gene therapy;
XX KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
XX KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
XX KW glucose transportation; catecholamine synthesis; iron transport;
XX KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;

DR WPI; 2004-070610/07.
 XX
 PT New antisense oligonucleotide hybridizable with a polynucleotide encoding
 PT a polypeptide with heparanase activity, useful for treating diseases such
 PT as cancer and autoimmune disorders.
 XX
 PS Claim 3; SEQ ID NO 10; 108pp; English.
 XX
 CC The invention relates to an antisense oligonucleotide (ASO) comprising a
 CC polynucleotide or a polynucleotide analogue of at least 10 bases being
 CC hybridisable in vivo, under physiological conditions, with a portion of
 CC a polynucleotide strand encoding a polypeptide having heparanase
 CC catalytic activity. Also included are a method of in vivo downregulating
 CC heparanase activity (comprising administering the ASO in vivo), a method
 CC of treating a subject suffering from a pathological condition
 CC (characterised by heparanase activity, comprising administering ASO to
 CC the subject), a pharmaceutical composition comprising the ASO and a
 CC carrier, an antisense nucleic acid construct (comprising a promoter
 CC sequence and a polynucleotide sequence directing the synthesis of an
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,
 CC under physiological conditions, with a polynucleotide strand encoding a
 CC polypeptide having heparanase catalytic activity), a method of in vivo
 CC downregulating heparanase activity (comprising administering in vivo the
 CC antisense nucleic acid construct), a pharmaceutical composition
 CC comprising the antisense nucleic acid construct and a carrier, and an
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
 CC analogue of at least 10 bases being hybridisable in vivo, under
 CC physiological conditions, with a portion of a polynucleotide strand being
 CC characterised by forming at least a portion of an untranslated region
 CC (UTR) for a polynucleotide strand encoding a polypeptide having
 CC heparanase catalytic activity. The methods and compositions of the
 CC present invention are useful for the prevention and/or treatment of
 CC diseases or conditions associated with aberrant heparanase activity, such
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and
 CC inflammation. The gene for human heparanase is located on chromosome 4.
 CC The present sequence is a human heparanase protein.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 104; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PAYLRFGGTKTDFLIFDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIFDPK 107
 |||||
 RESULT 40
 ID ADK52086
 XX ADK52086 standard; protein; 543 AA.
 AC ADK52086;
 XX
 DT 20-MAY-2004 (first entry)
 XX
 DE Human atopic dermatitis/psoriasis-associated protein #1.
 XX
 KW Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;
 KW antipsoriatic; rash.
 XX
 OS Homo sapiens.
 XX
 PN WO2004016785-A1.
 XX
 XX 26-FEB-2004.
 XX
 XX 06-AUG-2003; 2003WO-JP009999.
 PF
 XX
 PR 06-AUG-2002; 2002JP-00229319.
 PR
 PR 14-MAY-2003; 2003JP-00136544.
 XX
 XX (GENO-) GENOX RES INC.
 PA

PA (UYJU-) UNIV JUNTENDO.
 XX
 PI Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
 PI Mitsuishi K;
 XX
 DR WPI; 2004-214514/20.
 DR N-PSDB; ADK51968.
 XX
 PT Detecting atopic dermatitis or psoriasis comprises assaying levels of
 PT expression of an indicator gene at a rash site and non-rash site of a
 PT person with atopic dermatitis or psoriasis.
 XX
 PS Example 2; SEQ ID NO 119; 484pp; Japanese.
 XX
 CC The invention relates to detecting atopic dermatitis or psoriasis
 CC comprising assaying the levels of expression of an indicator gene at a
 CC rash site and non-rash site of a person with atopic dermatitis or
 CC psoriasis, comparing these levels with those of a healthy person, and
 CC determining that if the levels of indicators are higher or lower, then
 CC this indicates the disease. Also included are a reagent for detecting
 CC atopic dermatitis or psoriasis, a kit for screening for treatments, a
 CC transgenic non human vertebrate animal models for the diseases, an agent
 CC for inducing the diseases in mice and a DNA chip for assaying for the
 CC indicator genes. The method is used for treatment, detection and animal
 CC models for research of atopic dermatitis and psoriasis. The present
 CC sequence is a protein encoded by an indicator gene of the invention.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 104; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PAYLRFGGTKTDFLIFDPK 19
 |||||
 Db 89 PAYLRFGGTKTDFLIFDPK 107
 |||||
 RESULT 41
 ID ADM48716
 XX ADM48716 standard; protein; 543 AA.
 AC ADM48716;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human hpa protein #1.
 XX
 KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
 KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
 KW human.
 XX
 OS Homo sapiens.
 XX
 PN US2003217375-A1.
 XX
 PD 20-NOV-2003.
 XX
 XX 24-FEB-2003; 2003US-00371218.
 PF
 XX 31-AUG-1999; 98WO-US017954.
 PR
 PR 01-MAR-1999; 99US-00258892.
 PR
 PR 06-FEB-2001; 2001US-00776874.
 PR
 PR 19-NOV-2001; 2001US-00988113.
 XX
 XX (ZCHA/) ZCHARIA E.
 PA (VLOD/) VLODAVSKY I.
 PA (METZ/) METZGER S.
 PA (PECK/) PECKER I.
 PA (ILAN/) ILAN N.
 PA (CHAJ/) CHAJEK-SHAUL T.
 PA (GOLD/) GOLDSCHMIDT O.
 XX

PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX
DR WPI; 2004-021918/02.
DR N-PSDB; ADM48715, ADM48717.
XX
XX New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
PS Example 1; SEQ ID NO 10; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIIDPK 19
Db 89 PAYLRFGGTKTDFLIIDPK 107
RESULT 42
ADM48759
ID ADM48759 standard; protein; 543 AA.
XX
AC ADM48759;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human hpa protein #2.
XX
XX Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW human.
XX
XX Homo sapiens.
OS
XX US2003217375-A1.
PN
XX 20-NOV-2003.
PD
XX 24-FEB-2003; 2003US-00371218.
PF
XX 31-AUG-1998; 98WO-US017954.
PR
XX 01-MAR-1999; 99US-00258892.
PR
XX 06-FEB-2001; 2001US-00776874.
PR
XX 19-NOV-2001; 2001US-00988113.
XX
XX {ZCHA/} ZCHARIA E.
PA {VLOD/} VLODAVSKY I.
PA {METZ/} METZGER S.
PA {PECK/} PECKER I.
PA {ILAN/} ILAN N.
PA {CHAJ/} CHAJEK-SHAUL T.
PA {GOLD/} GOLDSHMIDT O.
XX
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;

XX WPI; 2004-021918/02.
DR N-PSDB; ADM48748.
XX
PT New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
PS Example 10; Fig 16; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIIDPK 19
Db 89 PAYLRFGGTKTDFLIIDPK 107
RESULT 43
ADM05074
ID ADM05074 standard; protein; 543 AA.
XX
AC ADM05074;
XX
XX 01-JUL-2004 (first entry)
XX
XX Antipsoriatic protein sequence #716.
XX
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
OS
XX WO2004028479-A2.
PN
XX 08-APR-2004.
PD
XX 25-SEP-2003; 2003WO-US030907.
PF
XX 25-SEP-2002; 2002US-0414006P.
PR
XX (GETH) GENENTECH INC.
PA
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
PI
XX WPI; 2004-305105/28.
DR N-PSDB; ADM05073.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
PS Claim 9; SEQ ID NO 1468; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition

CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19
|||||
Db 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 44

ADN04902
ID ADN04902 standard; protein; 543 AA.

XX AC

ADN04902;

XX DT 01-JUL-2004 (first entry)

XX DE Antipsoriatic protein sequence #631.

XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.

XX OS Homo sapiens.

XX PN WO2004028479-A2.

XX PD 08-APR-2004.

XX PF 25-SEP-2003; 2003WO-US030907.

XX PR 25-SEP-2002; 2002US-0414006P.

XX PA (GETH) GENENTECH INC.

XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;

XX DR WPI; 2004-305105/28.

XX DR N-PSDB; ADN04901.

XX PT New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.

XX PS Claim 9; SEQ ID NO 1296; 3069pp; English.

XX CC The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19

|||||

Db 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 45

AD063831

ID AD063831 standard; protein; 543 AA.

XX AC

AD063831;

XX

DT 26-AUG-2004 (first entry)

DE Human heparanase mutant E378A.

XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein; enzyme.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Active-site 225

FT /note= "Active site proton donor"

FT Active-site 343

FT /note= "Active site nucleophile"

FT Misc-difference 378

FT /note= "Ala replaces wild-type Glu"

XX WO2004048558-A2.

XX PD 10-JUN-2004.

XX PF 24-NOV-2003; 2003WO-IL000989.

XX PR 24-NOV-2002; 2002IL-00153059.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

XX WPI; 2004-450373/42.

XX PT New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.

XX PS Example 4; Page; 128pp; English.

XX CC The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (AD063822-AD063824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (AD063825-
CC AD063827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a human
CC heparanase mutant E378A created in an example of the invention which
CC retains its heparanase catalytic activity. The present sequence is not
CC shown in the invention, but is derived from the protein sequence of
CC GenBank accession number AF144325 and the information provided on page
CC 70.

CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents the human
CC CC heparanase double mutant E225A/E343A.
XX XX

SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGKTDFLIIDPK 19
 |||||
DB 89 PAYLRFGGKTDFLIIDPK 107
 |||||

RESULT 47

ADO63823

ID ADO63823 standard; protein; 543 AA.

XX AC ADO63823;

XX DT 26-AUG-2004 (first entry)

XX DE Human heparanase mutant E343A, SEQ ID: 8.

KW Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnary; mutant; mutein.
XX OS Homo sapiens.
OS Synthetic.

FH Key Location/Qualifiers
FT Active-site 225 /note= "Active site proton donor"
FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site
FT nucleophile)"

XX WO2004048558-A2.
XX 10-JUN-2004.
XX 24-NOV-2003; 2003WO-IL000989.
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX N-PSDB; ADO63817.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.

XX PS Claim 9; SEQ ID NO 8; 128pp; English.

XX PS

CC The invention relates to nucleic acid constructs comprising a nucleic acid encoding a heparanase-derived protein which lacks heparanase endoglycosidase catalytic activity but which retains its cell-cell and cell-matrix adhesion properties. The constructs of the invention optionally further comprise operably linked regulatory elements. The invention also relates to the heparanase-derived proteins and host cells comprising the nucleic acid constructs of the invention. The heparanase-derived proteins are especially mutants of human heparanase in which the active site proton donor Glu225 and/or the active site nucleophile Glu343 are replaced with Ala (ADO63822-ADO63824), and the proteins may optionally further comprise an avian heparanase signal peptide (ADO63825-ADO63827). The heparanase-derived protein, nucleic acid construct and host cells are useful in preparing a tissue sealant composition for sealing injuries, reducing the loss of blood, accelerating the healing and homeostasis of an injury, accelerating blood vessel endothelium formation or the endothelialisation of vascular grafts, accelerating the adhesive activity of mammalian cells, and accelerating the adhesion and aggregation of platelets. They may also be used in the treatment of disorders associated with adhesion deficiency such as LAD (leukocyte adhesion deficiency), Glanzmann's thrombasthenia (defective platelet function), or Bernard-Soulier syndrome (deficient platelet adhesion). The cells of the invention may additionally be used to screen for modulators of cell-cell and cell-matrix adhesion, and to prepare an implantable synthetic vascular graft comprising a tube made of a biocompatible material lined with the cells. The present sequence represents the human heparanase mutant E343A.

XX PS Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19
Db |||||

ADO63822

AD ADO63832 standard; protein; 543 AA.

AC ADO63832;

DT 26-AUG-2004 (first entry)

XX Human heparanase mutant E396A.

DE

XX Human; heparanase; heparanase-derived protein; heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion; tissue sealant; injury; blood loss; endothelialisation; blood vessel; vascular graft; platelet adhesion; platelet aggregation; adhesion disorder; LAD; leukocyte adhesion deficiency; Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening; vulnary; mutant; mutein; enzyme.

XX Homo sapiens.

OS Synthetic.

XX

FX Key Location/Qualifiers

FT Active-site 225

FT Active-site /note= "Active site proton donor"

FT Active-site 343

FT /note= "Active site nucleophile"

FT Misc-difference 396

FT /note= "Ala replaces wild-type Glu"

FT

XX WO2004048558-A2.

PN

XX 10-JUN-2004.

PD

XX 24-NOV-2003; 2003WO-IL000989.

XX 24-NOV-2002; 2002IL-00153059.

PR (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

PI WPI; 2004-450373/42.

DR

XX New nucleic acid construct comprising heparanase-derived polypeptide, useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's thrombasthenia, or Bernard-Soulier syndrome.

PT

XX Example 4; Page; 128pp; English.

PS The invention relates to nucleic acid constructs comprising a nucleic acid encoding a heparanase-derived protein which lacks heparanase endoglycosidase catalytic activity but which retains its cell-cell and cell-matrix adhesion properties. The constructs of the invention optionally further comprise operably linked regulatory elements. The invention also relates to the heparanase-derived proteins and host cells comprising the nucleic acid constructs of the invention. The heparanase-derived proteins are especially mutants of human heparanase in which the active site proton donor Glu225 and/or the active site nucleophile Glu343 are replaced with Ala (ADO63822-ADO63824), and the proteins may optionally further comprise an avian heparanase signal peptide (ADO63825-ADO63827). The heparanase-derived protein, nucleic acid construct and host cells are useful in preparing a tissue sealant composition for sealing injuries, reducing the loss of blood, accelerating the healing and homeostasis of an injury, accelerating blood vessel endothelium formation or the endothelialisation of vascular grafts, accelerating the adhesive activity of mammalian cells, and accelerating the adhesion and aggregation of platelets. They may also be used in the treatment of disorders associated with adhesion deficiency such as LAD (leukocyte adhesion deficiency), Glanzmann's thrombasthenia (defective platelet function), or Bernard-Soulier syndrome (deficient platelet adhesion). The cells of the invention may additionally be used to screen for modulators of cell-cell and cell-matrix adhesion, and to prepare an implantable synthetic vascular graft comprising a tube made of a biocompatible material lined with the cells. The present sequence represents a human heparanase mutant E378A created in an example of the invention which retains its heparanase catalytic activity. The present sequence is not shown in the invention, but is derived from the protein sequence of GenBank accession number AF144325 and the information provided on page 70.

XX Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19
Db |||||

ADO63822

AD ADO63822 standard; protein; 543 AA.

AC ADO63822;

DT 26-AUG-2004 (first entry)

XX Human heparanase mutant E225A, SEQ ID:7.

DE

XX Human; heparanase; heparanase-derived protein; heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion; tissue sealant; injury; blood loss; endothelialisation; blood vessel; vascular graft; platelet adhesion; platelet aggregation;

XX

KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
XX vulnary; mutant; mutein.
OS Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 225
FT /note= "Ala replaces wild-type Glu (active site proton
FT donor)"
FT Active-site 343
FT /note= "Active site nucleophile"
XX WO2004048558-A2.
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-1L000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX
XX WPI: 2004-450373/42.
XX N-PSDB; ADO63816.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Claim 9; SEQ ID NO 7; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be use in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents the human
CC heparanase mutant E225A.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PAYLRFGGTKTDFLIFDPK 19
Db 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 50
ADQ80372
ID ADQ80372 standard; protein; 543 AA.
XX
XX AC ADQ80372;
XX DT 21-OCT-2004 (first entry)
XX DE Heparanase protein.
XX KW cytosatic; epidermal growth factor receptor modulator; identification;
KW therapeutic response; cancer; EGFR; biomarker.
XX OS Homo sapiens.
XX PN WO2004063709-A2.
XX PD 29-JUL-2004.
XX PF 08-JAN-2004; 2004WO-US000368.
XX PR 08-JAN-2003; 2003US-0438735P.
XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX PI Amler LC, Januario T;
XX WPI: 2004-544114/52.
XX N-PSDB; ADQ80253.
XX
XX Identifying a mammal that will respond therapeutically to a method of
PT treating cancer comprises comparing the level of a biomarker in a mammal
PT before and after exposure to an epidermal growth factor receptor (EGFR)
PT modulator.
XX
XX Disclosure; SEQ ID NO 144; 520pp; English.
XX
XX The invention relates to a method of identifying a mammal that will
CC respond therapeutically to a method of treating cancer by administering
CC an epidermal growth factor receptor (EGFR) modulator by comparing the
CC level of a biomarker in a mammal before and after exposure to an EGFR
CC modulator. The method comprises: (a) measuring, in the mammal, the level
CC of at least one biomarker identified in the specification; (b) exposing
CC the mammal to the EGFR modulator; and (c) measuring in the mammal the
CC level of the biomarker where a difference in the level in step (c)
CC compared to step (a) indicates that the mammal will respond
CC therapeutically to the method of treating cancer. The method and
CC biomarkers are useful for identifying a mammal that will respond
CC therapeutically to a method of treating cancer by administering an
CC epidermal growth factor receptor (EGFR) modulator. This sequence
CC corresponds to one of the biomarkers whose levels of expression is
CC measured in the method of the invention.
XX
SQ Sequence 543 AA;
Query Match 100.0%; Score 104; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 6.9e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PAYLRFGGTKTDFLIFDPK 19
Db 89 PAYLRFGGTKTDFLIFDPK 107

Search completed: June 5, 2006, 12:41:31
Job time : 162.219 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 21.0822 Seconds
(without alignments)
86.714 Million cell updates/sec

Title: US-10-645-659A-7

Perfect score: 104

Sequence: 1 PAYLRFGGKTDFLIFDPK 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: PIR-80.*

2: pir1.*

3: pir2.*

4: pir3.*

5: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	63.5	480	2 D73506	heparanase protein
2	49	47.1	588	2 D70394	mannosyltransferase
3	48	46.2	264	1 CRH07	carbonate dehydrat
4	48	46.2	265	2 C86516	hypothetical prote
5	48	46.2	265	2 D72105	KDO-transferase 2,
6	47	45.2	176	2 E72351	hypothetical prote
7	46	44.2	405	2 E30986	probable colanic a
8	46	44.2	405	2 H85831	Probable colanic a
9	46	44.2	405	2 G84971	Putative colanic a
10	45	43.3	241	2 S77596	cytochrome-c oxida
11	45	43.3	246	1 B55582	cytochrome-c oxida
12	44	42.3	284	2 F84363	Brp-like homolog [
13	44	42.3	762	2 D87403	cbbC protein [imp
14	44	42.3	997	2 JC6067	CCAAT-binding fac
15	44	42.3	998	2 A36368	transcription fac
16	43	41.3	242	1 S49346	cytochrome-c oxida
17	43	41.3	244	1 B47468	cytochrome-c oxida
18	43	41.3	253	2 B71346	probable SpoOJ reg
19	43	41.3	1293	2 T01512	hypothetical prote
20	42.5	40.9	175	2 T21297	FMRFamide-like neu
21	42	40.4	219	2 T26632	hypothetical prote
22	42	40.4	240	2 A69191	tRNA nucleotidyltr
23	42	40.4	241	2 H81327	hypothetical prote
24	42	40.4	687	2 H72485	probable hydantoin
25	42	40.4	742	2 A49672	transcription fac
26	42	40.4	772	2 A55004	transcription fac
27	42	40.4	857	2 JC4169	phosphoenolpyruvat
28	41	39.4	144	2 T24474	hypothetical prote
29	41	39.4	227	2 S76494	hypothetical prote

30	41	39.4	262	1 S76610	hypothetical prote
31	41	39.4	444	2 A72209	hypothetical prote
32	41	39.4	471	2 A86303	hypothetical prote
33	41	39.4	660	2 A70362	N-methylhydantoina
34	41	39.4	785	2 B72608	probable hyuA APE1
35	41	39.4	5069	2 T17464	rifamycin polyketi
36	40.5	38.9	164	2 B44827	FMRFamide-like pep
37	40	38.5	99	2 G98278	hypothetical prote
38	40	38.5	141	2 T51959	hypothetical prote
39	40	38.5	178	2 AD1156	hypothetical prote
40	40	38.5	221	2 AB2204	lipote-protein li
41	40	38.5	243	2 AG3447	cytochrome-c oxida
42	40	38.5	286	2 H97341	dihydrodipicolinat
43	40	38.5	310	2 T35417	probable beta-lact
44	40	38.5	338	2 D89102	protein F25E5.2 li
45	40	38.5	346	2 AC0382	probable Pyridoxal
46	40	38.5	349	2 T03500	hypothetical prote
47	40	38.5	381	1 A47327	selenoprotein P pr
48	40	38.5	396	2 S56954	protein YBR162c ho
49	40	38.5	419	2 T20514	hypothetical prote
50	40	38.5	433	2 G71657	polyglutamate
51	40	38.5	438	2 B97712	hypothetical prote
52	40	38.5	528	2 S38242	hypothetical prote
53	40	38.5	576	2 T24193	hypothetical prote
54	40	38.5	585	2 T18885	hypothetical prote
55	40	38.5	742	2 A49340	alcohol dehydrogen
56	40	38.5	742	2 JS0326	alcohol dehydrogen
57	40	38.5	762	2 T00410	protein kinase hom
58	40	38.5	967	2 H86334	T20H2.10 protein -
59	39.5	38.0	223	2 B81378	two-component regu
60	39.5	38.0	246	2 T44624	cini protein [impo
61	39.5	38.0	289	2 T35113	diaminopimelate ep
62	39.5	38.0	481	2 B88086	protein Tlflr1.2 li
63	39	37.5	143	2 T24261	hypothetical prote
64	39	37.5	158	2 AE0444	hypothetical prote
65	39	37.5	186	2 C87500	conserved hypotet
66	39	37.5	192	2 A64098	molybdopterin-guan
67	39	37.5	243	1 S39989	cytochrome-c oxida
68	39	37.5	243	2 AG2765	cytochrome C oxida
69	39	37.5	243	2 D97546	cytochrome-c oxida
70	39	37.5	243	2 C95313	cytochrome-c oxida
71	39	37.5	243	2 H95344	FixO1 c-type cytoc
72	39	37.5	264	2 F83351	conserved hypotet
73	39	37.5	296	2 A64035	hypothetical prote
74	39	37.5	301	2 B71246	hypothetical prote
75	39	37.5	319	2 T45980	hypothetical prote
76	39	37.5	324	2 H81553	probable transcrip
77	39	37.5	329	2 H71192	hypothetical prote
78	39	37.5	339	2 A60918	FMRFamide precurs
79	39	37.5	380	2 T10442	selenoprotein P pr
80	39	37.5	385	1 OMRTSP	selenoprotein P pr
81	39	37.5	404	2 S57178	2-nitropropane dio
82	39	37.5	422	2 C70518	probable nant prot
83	39	37.5	425	2 AH1041	Vi polysaccharide
84	39	37.5	425	2 B36892	Vi polysaccharide
85	39	37.5	433	2 S75948	hypothetical prote
86	39	37.5	454	2 A43501	sucrose-6-phosphat
87	39	37.5	517	2 C69550	conserved hypotet
88	39	37.5	580	2 A83874	carbon starvation-
89	39	37.5	581	2 T50841	phosphinositide-s
90	39	37.5	661	2 A32475	dnak-type molecula
91	39	37.5	661	2 T15513	heat shock 70K pro
92	39	37.5	831	2 AB3513	ATPase virB4 homol
93	39	37.5	849	2 D82785	mannosyltransferas
94	39	37.5	898	2 H87481	ribonuclease, Rne/
95	39	37.5	1068	2 T04112	pol protein homolo
96	38.5	37.0	400	2 E75135	methylmalonyl-coa
97	38.5	37.0	400	2 H71073	probable methylmal
98	38.5	37.0	695	2 T40168	hypothetical prote
99	38.5	37.0	984	2 T48216	hypothetical prote
100	38.5	37.0	1545	2 T14288	DNA (cytosine-5-)-
101	38.5	37.0	1761	2 T14289	DNA (cytosine-5-)-
102	38	36.5	134	2 F36789	hypothetical prote

103	38	36.5	134	2	AC0079	hypothetical prote	176	37	35.6	330	2	T35886	hypothetical prote
104	38	36.5	198	2	S60969	probable membrane	177	37	35.6	336	2	B71121	hypothetical prote
105	38	36.5	217	2	F87372	peptide methionine	178	37	35.6	350	2	AE3022	threonine aldolase
106	38	36.5	303	2	T29321	hypothetical prote	179	37	35.6	350	2	E98262	hypothetical prote
107	38	36.5	313	2	H64144	hypothetical prote	180	37	35.6	356	2	B90271	aspartate aminotra
108	38	36.5	313	2	C85036	hypothetical prote	181	37	35.6	382	2	S51962	FUN49 protein - ye
109	38	36.5	338	2	E72334	conserved hypothet	182	37	35.6	389	2	T44248	sarcosine oxidase
110	38	36.5	341	1	DEK601	glyceroldehyde-3-p	183	37	35.6	390	2	T39975	sarcosine oxidase
111	38	36.5	341	1	DEK604	glyceroldehyde-3-p	184	37	35.6	402	2	D82980	hypothetical prote
112	38	36.5	346	2	G86305	SRG1 homolog [impo	185	37	35.6	432	2	C81308	UDP-N-acetylmurama
113	38	36.5	359	2	T59403	alpha-2, 8-polysial	186	37	35.6	438	2	A70544	probable histidino
114	38	36.5	359	2	S52425	polysialyltransfer	187	37	35.6	439	2	G97159	contains cell adhe
115	38	36.5	359	2	JC4224	alpha-N-acetylneur	188	37	35.6	440	2	T11319	NADH2 dehydrogenas
116	38	36.5	371	2	T21153	hypothetical prote	189	37	35.6	453	2	C83722	cytochrome P450 BH
117	38	36.5	379	1	DWDXBF	2-hydroxyglutaryl-	190	37	35.6	470	2	G70564	probable membrane
118	38	36.5	409	2	C95041	hypothetical prote	191	37	35.6	473	2	A84979	sulfate adenyllytr
119	38	36.5	409	2	A86240	protein F20824.10	192	37	35.6	516	2	AC1132	hydantoinase homol
120	38	36.5	414	2	A87165	L-lactate dehydrog	193	37	35.6	516	2	AC1132	hydantoinase homol
121	38	36.5	426	2	C97797	tetrahydrofolylpol	194	37	35.6	529	2	F69020	conserved hypothet
122	38	36.5	440	2	F83235	outer membrane por	195	37	35.6	537	2	B46535	interleukin 2 rece
123	38	36.5	440	2	S11793	porin P precursor,	196	37	35.6	565	2	T47423	hypothetical prote
124	38	36.5	440	2	D71715	hypothetical prote	197	37	35.6	565	2	T29813	hypothetical prote
125	38	36.5	463	2	H72365	heat shock protein	198	37	35.6	578	2	B89045	protein B0238.7 li
126	38	36.5	464	2	B35159	arylsulfatase (EC	199	37	35.6	607	2	A26168	riboophorin I precu
127	38	36.5	521	2	T45608	hypothetical prote	200	37	35.6	607	2	S60905	hypothetical prote
128	38	36.5	553	2	S75892	probable glycerol-	201	37	35.6	639	2	C42049	leishmanolysin (EC
129	38	36.5	577	2	T45548	arylsulfatase (EC	202	37	35.6	644	2	E90367	n-methyl hydantoin
130	38	36.5	608	2	T32708	hypothetical prote	203	37	35.6	645	2	T49702	related to DOS1 pr
131	38	36.5	634	2	T25425	hypothetical prote	204	37	35.6	646	1	S19916	leishmanolysin (EC
132	38	36.5	725	2	D87002	probable cation tr	205	37	35.6	649	2	T33431	hypothetical prote
133	38	36.5	728	2	S55601	hypothetical prote	206	37	35.6	680	2	D64420	N-methylhydantoina
134	38	36.5	738	2	S14270	alcohol dehydrogen	207	37	35.6	684	2	T45534	agaf protein [impo
135	38	36.5	741	2	T12706	NADH2 dehydrogenas	208	37	35.6	787	2	C48898	hypothetical prote
136	38	36.5	834	2	S66498	M-sema F protein p	209	37	35.6	853	2	AC2079	ferrichrome-iron r
137	38	36.5	924	2	C66725	chromosome segrega	210	37	35.6	879	2	S49910	chloroplast outer
138	38	36.5	976	2	PC4208	valine-tRNA ligase	211	37	35.6	884	2	S66308	nitrate reductase
139	38	36.5	1048	2	T31653	hypothetical prote	212	37	35.6	891	2	AC3384	ribonuclease E / z
140	38	36.5	1163	2	T24855	hypothetical prote	213	37	35.6	936	2	D97630	hypothetical prote
141	38	36.5	1175	2	S39951	chitin synthase (E	214	37	35.6	1025	2	G81722	polymorphic membra
142	38	36.5	1205	2	AH2486	hypothetical prote	215	37	35.6	1289	2	S69689	hypothetical prote
143	38	36.5	1268	2	B36502	insulin receptor-r	216	37	35.6	1873	2	A30063	dihydropyridine re
144	38	36.5	1339	1	S20052	DNA-directed DNA p	217	37	35.6	2332	1	GNNYF	genome polyprotein
145	38	36.5	1374	2	T16129	hypothetical prote	218	37	35.6	3083	2	AH2493	hypothetical prote
146	38	36.5	2100	2	T03223	probable polyketid	219	36.5	35.1	91	2	S25462	Ig kappa chain V r
147	38	36.5	2240	2	T37057	probable multi-dom	220	36.5	35.1	95	2	S25177	Ig kappa chain V r
148	37.5	36.1	214	2	A87276	hypothetical prote	221	36.5	35.1	111	2	E53285	Ig kappa chain V a
149	37.5	36.1	347	2	A34616	FMRFamide polyprot	222	36.5	35.1	111	2	A33936	Ig kappa chain V r
150	37.5	36.1	362	2	G71167	hypothetical prote	223	36.5	35.1	131	1	KVNSM6	Ig kappa chain pre
151	37.5	36.1	375	2	JC6169	nuclear IIM intera	224	36.5	35.1	143	2	S77214	hypothetical prote
152	37.5	36.1	386	2	T04914	hypothetical prote	225	36.5	35.1	269	2	D90141	hypothetical prote
153	37.5	36.1	430	2	T28143	tapasin I homolog,	226	36.5	35.1	280	2	T35432	probable aldehyde
154	37.5	36.1	566	2	T33042	hypothetical prote	227	36.5	35.1	313	2	B97425	flaC protein (X964
155	37.5	36.1	705	2	T51034	hypothetical prote	228	36.5	35.1	313	2	A12642	flagellin [importe
156	37.5	36.1	1213	2	T51032	hypothetical prote	229	36.5	35.1	315	2	T35804	probable aldehyde
157	37.5	36.1	1456	2	T15961	hypothetical prote	230	36.5	35.1	448	2	F83628	beta-alanine-pyruv
158	37.5	36.1	1559	2	T07757	probable DNA (cyto	231	36.5	35.1	449	1	A42800	beta-alanine-pyruv
159	37.5	36.1	1699	2	T31340	voltage-gated sodi	232	36.5	35.1	464	2	H70169	phosphogluconate d
160	37.5	36.1	4588	2	T28667	dyein beta heavy	233	36.5	35.1	482	2	D83379	probable aldehyde
161	37	35.6	99	2	S28707	hypothetical prote	234	36.5	35.1	483	2	H98159	hypothetical prote
162	37	35.6	135	2	T21607	FMRFamide-like pep	235	36.5	35.1	483	2	A13127	vanillin: NAD oxid
163	37	35.6	198	2	T48931	hypothetical prote	236	36.5	35.1	545	1	S08020	RNA-directed RNA p
164	37	35.6	202	2	A64480	hypothetical prote	237	36.5	35.1	600	2	B84948	NADH2 dehydrogenas
165	37	35.6	224	2	G82913	conserved hypothet	238	36.5	35.1	653	2	F70383	organic solvent to
166	37	35.6	228	2	C69859	two-component resp	239	36.5	35.1	1400	2	A81672	secDF protein, pro
167	37	35.6	254	2	B84901	hypothetical prote	240	36.5	35.1	1554	2	T63370	probable DNA (cyto
168	37	35.6	275	2	E82554	extragenic suppress	241	36	34.6	70	2	S01213	NADH2 dehydrogenas
169	37	35.6	292	2	B82411	transcription regu	242	36	34.6	81	2	C86710	hypothetical prote
170	37	35.6	300	2	A82968	hypothetical prote	243	36	34.6	105	2	AC0277	probable phage pro
171	37	35.6	300	2	E98314	probable permease	244	36	34.6	107	2	A75089	hypothetical prote
172	37	35.6	315	2	T26673	hypothetical prote	245	36	34.6	115	2	AF1366	transcription regu
173	37	35.6	326	2	F82334	hflC protein VC035	246	36	34.6	115	2	AG1735	transcription regu
174	37	35.6	328	2	T16065	iron-sulfur cofact	247	36	34.6	126	2	T16727	hypothetical prote
175	37	35.6	330	2	A70422	hypothetical prote	248	36	34.6	137	2	T03491	conserved hypothet

249	36	34.6	143	2	A83328	probable ring-clea	322	36	34.6	634	2	S32349	probable SNF2-type
250	36	34.6	155	2	C97885	hypothetical prote	323	36	34.6	640	2	F90587	lipoprotein limpor
251	36	34.6	156	2	JQ0351	heat shock protein	324	36	34.6	650	2	A90473	n-methylhydantoina
252	36	34.6	156	2	T49264	heat shock protein	325	36	34.6	664	2	T39959	probable urea acti
253	36	34.6	157	2	S06074	heat shock protein	326	36	34.6	666	2	B70320	nitrate reductase
254	36	34.6	157	2	T14381	heat-shock protein	327	36	34.6	686	2	T10684	hypothetical prote
255	36	34.6	173	1	RUPSE0	rubredoxin II - Ps	328	36	34.6	722	1	VCPVCN	coat protein VPI -
256	36	34.6	206	2	C83026	conserved hypotet	329	36	34.6	722	1	VCPVME	coat protein VPI -
257	36	34.6	224	2	F64339	hypothetical prote	330	36	34.6	727	1	VCPVIF	coat protein VPI -
258	36	34.6	229	2	I50106	MHC class I histoc	331	36	34.6	727	1	VCPVFP	coat protein VPI -
259	36	34.6	239	2	F95374	hypothetical prote	332	36	34.6	735	2	I48101	ADAM 6 protein pre
260	36	34.6	244	2	S94911	hypothetical prote	333	36	34.6	737	1	VCPVCD	coat protein VPI -
261	36	34.6	247	2	T11315	ATP synthase chain	334	36	34.6	741	2	I48694	probable transcrip
262	36	34.6	256	1	H70023	N-acetyl-glucosami	335	36	34.6	747	2	T34329	hypothetical prote
263	36	34.6	260	2	T08463	carbonate dehydrat	336	36	34.6	748	1	VCPVCP	coat protein VPI -
264	36	34.6	262	2	S06060	gene NDI intron 4	337	36	34.6	749	2	G64165	ATP-dependent heli
265	36	34.6	266	2	C95072	hypothetical prote	338	36	34.6	812	2	AC2349	hypothetical prote
266	36	34.6	271	2	A97940	hypothetical prote	339	36	34.6	819	2	T22152	hypothetical prote
267	36	34.6	280	1	RLBH	rRNA N-glycosidase	340	36	34.6	820	2	A65247	hypothetical prote
268	36	34.6	281	2	B38664	30K ribosome inact	341	36	34.6	837	2	T19271	hypothetical prote
269	36	34.6	295	2	J64744	NAD-dinitrogen-red	342	36	34.6	844	1	A28528	penicillin amidase
270	36	34.6	295	2	I39751	NAD-dinitrogen-red	343	36	34.6	846	1	PNECA	penicillin amidase
271	36	34.6	298	2	T17568	hydrolase homolog	344	36	34.6	858	2	T47223	replication licens
272	36	34.6	318	2	C70661	probable moew - My	345	36	34.6	862	1	S56766	replication licens
273	36	34.6	320	2	T09555	fibrillarlin - Arab	346	36	34.6	863	1	S64720	replication licens
274	36	34.6	328	2	D84073	tellurite resistan	347	36	34.6	863	1	S65954	replication licens
275	36	34.6	336	2	A96997	ferrichrome transp	348	36	34.6	871	2	T40845	dna ligase - flesi
276	36	34.6	342	2	S76447	hypothetical prote	349	36	34.6	873	2	H96503	protein F9C16.17 [
277	36	34.6	343	2	A53057	retinal-binding pr	350	36	34.6	877	2	H71647	alanine-tRNA ligas
278	36	34.6	351	2	AF0559	probable lyase STY	351	36	34.6	878	2	G97865	alanine-tRNA ligas
279	36	34.6	370	2	C70464	GTP-binding protei	352	36	34.6	917	2	F95884	probable sensory h
280	36	34.6	373	2	F83020	UDP-glucose-heptos	353	36	34.6	969	2	F71418	hypothetical prote
281	36	34.6	374	2	AC0972	lipopolysaccharide	354	36	34.6	1070	2	S19686	alpha-glucosidase
282	36	34.6	374	2	B42595	glucosyltransferas	355	36	34.6	1116	2	T38073	serine/threonine-p
283	36	34.6	374	2	B98192	glucosyltransferas	356	36	34.6	1255	2	T31065	diaphanous protein
284	36	34.6	374	2	C96039	hypothetical prote	357	36	34.6	1260	2	T37523	probable oxoprolin
285	36	34.6	389	2	D81139	hypothetical prote	358	36	34.6	1483	2	T19751	hypothetical prote
286	36	34.6	401	2	S53862	NADH2 dehydrogenas	359	36	34.6	1517	2	T38912	hypothetical integ
287	36	34.6	404	2	T44590	tylosin biosynthes	360	36	34.6	1618	2	S21424	nestin - human
288	36	34.6	409	2	T11901	NADH2 dehydrogenas	361	36	34.6	1755	2	F82618	chemotaxis-related
289	36	34.6	415	2	T23215	hypothetical prote	362	36	34.6	2139	2	A44467	voltage-dependent
290	36	34.6	416	2	F70593	hypothetical prote	363	36	34.6	2143	2	JH0427	voltage-dependent
291	36	34.6	418	2	D81932	NADH2 dehydrogenas	364	36	34.6	2163	2	S50675	pre-mRNA splicing
292	36	34.6	418	2	B81222	NADH2 dehydrogenase	365	36	34.6	2166	2	S11339	calcium channel pr
293	36	34.6	424	2	H96963	dihydroorotase [im	366	36	34.6	2171	2	S05054	calcium channel al
294	36	34.6	426	2	T51373	hypothetical prote	367	36	34.6	2220	2	A45290	calcium channel pr
295	36	34.6	428	2	D83861	hypothetical prote	368	36	34.6	3491	2	T43231	probable 6-deoxyer
296	36	34.6	432	2	C54088	trigger factor tig	369	36	34.6	3573	2	S23070	erythronolide synt
297	36	34.6	438	2	A72430	hypothetical prote	370	36	34.6	3871	2	T22812	hypothetical prote
298	36	34.6	445	1	S26209	histidinol dehydro	371	35.5	34.1	65	2	B75298	hypothetical prote
299	36	34.6	452	2	T25076	hypothetical prote	372	35.5	34.1	73	2	A13065	hypothetical prote
300	36	34.6	454	2	A69763	homoserine dehydro	373	35.5	34.1	73	2	G98220	hypothetical prote
301	36	34.6	456	2	D70772	hypothetical prote	374	35.5	34.1	194	2	D69486	probable ribosomal
302	36	34.6	459	2	B44498	radial spoke prote	375	35.5	34.1	233	2	A13590	DNA-directed DNA p
303	36	34.6	474	2	H90582	preprotein translo	376	35.5	34.1	294	2	G82972	conserved hypotet
304	36	34.6	479	2	S68598	sucrose-6-phosphat	377	35.5	34.1	357	2	F75066	hypothetical prote
305	36	34.6	492	2	F64464	sodium-dependent n	378	35.5	34.1	378	1	T03758	probable ferredoxi
306	36	34.6	507	1	A32966	cytochrome P450 4A	379	35.5	34.1	378	2	A47300	cell adhesion prot
307	36	34.6	507	2	T01555	hypothetical prote	380	35.5	34.1	383	2	F89916	hippurate hydrolas
308	36	34.6	514	2	T44976	hydantoïnase homol	381	35.5	34.1	383	2	T23057	hypothetical prote
309	36	34.6	519	2	J05315	dihydropyrimidin	382	35.5	34.1	387	2	S56411	hypothetical 45K p
310	36	34.6	522	2	S33029	hypothetical prote	383	35.5	34.1	414	2	I39840	hypothetical prote
311	36	34.6	522	2	T21591	hypothetical prote	384	35.5	34.1	423	2	B83106	hypothetical prote
312	36	34.6	531	2	C75418	ribonucleoprotein	385	35.5	34.1	424	2	B56144	aspartate carbamoy
313	36	34.6	556	2	I61581	transcription fact	386	35.5	34.1	455	2	A0163	glycylpeptide N-te
314	36	34.6	590	1	A45621	leishmanolysin (EC	387	35.5	34.1	484	2	AG3184	aldenhyde dehydrog
315	36	34.6	599	2	B42049	leishmanolysin (EC	388	35.5	34.1	484	2	A64251	glutamate-tRNA lig
316	36	34.6	599	2	A44951	leishmanolysin (EC	389	35.5	34.1	486	2	T26483	hypothetical prote
317	36	34.6	603	2	G84554	probable acyl-CoA	390	35.5	34.1	508	1	A30007	dolichyl-diphospho
318	36	34.6	605	2	A27274	ribofhorin i precu	391	35.5	34.1	508	1	ISHUSS	protein disulfide-
319	36	34.6	614	2	F96791	hypothetical prote	392	35.5	34.1	508	1	ISRTSS	protein disulfide-
320	36	34.6	628	2	H89917	conserved hypotet	393	35.5	34.1	509	1	ISMSSS	protein disulfide-
321	36	34.6	634	2	I40217	glutaryl 7-ACA acy	394	35.5	34.1	510	1	ISBOSS	protein disulfide-

395	35.5	34.1	515	1	ISCHS5	protein disulfide-	468	35	33.7	378	2	S77445	3-amino-5-hydroxyb
396	35.5	34.1	598	2	AG0311	NADH2 dehydrogenas	469	35	33.7	388	2	G90450	hypothetical prote
397	35.5	34.1	735	2	F81993	DNA helicase II (E	470	35	33.7	390	1	RREPT4	recombination/rep
398	35.5	34.1	963	2	T04002	hypothetical prote	471	35	33.7	393	2	B72204	maltose ABC transp
399	35.5	34.1	980	2	T38632	probable phosphati	472	35	33.7	394	2	AG3616	hypothetical prote
400	35.5	34.1	1749	2	S75138	hypothetical prote	473	35	33.7	401	1	B35177	chromate resistanc
401	35	33.7	30	2	E31461	T-cell receptor de	474	35	33.7	409	1	S75020	acetyl-CoA C-acety
402	35	33.7	147	2	F71089	N1,Fe-Hydrogenase	475	35	33.7	411	2	G97802	tyrosine-tRNA liga
403	35	33.7	156	2	A69966	hypothetical prote	476	35	33.7	412	2	T45097	hypothetical prote
404	35	33.7	156	2	A32868	prolactin receptor	477	35	33.7	412	2	A41070	prolactin receptor
405	35	33.7	158	2	F90127	hypothetical prote	478	35	33.7	419	2	H90480	conserved hypothet
406	35	33.7	162	2	G70232	hypothetical prote	479	35	33.7	426	2	T16406	hypothetical prote
407	35	33.7	168	2	E87709	hypothetical prote	480	35	33.7	428	2	H97118	levansucrase limpo
408	35	33.7	173	1	NKVGCU	core protein p20 -	481	35	33.7	432	2	D97851	UDP-glucose 6-dehy
409	35	33.7	173	2	A26659	T-cell receptor ga	482	35	33.7	433	2	A70536	probable pepC prot
410	35	33.7	179	2	D72360	conserved hypothet	483	35	33.7	434	2	E71638	UDP-glucose 6-dehy
411	35	33.7	195	2	D69505	phosphatidylserine	484	35	33.7	437	2	T44520	lipopolysaccharide
412	35	33.7	215	2	AB3358	probable carnitine	485	35	33.7	437	2	T44509	Vi polysaccharide
413	35	33.7	217	2	T27524	hypothetical prote	486	35	33.7	442	2	T34714	hypothetical prote
414	35	33.7	218	2	D96012	hypothetical membr	487	35	33.7	454	2	JC7231	thermophilic desul
415	35	33.7	224	2	AF3382	alpha/beta hydrola	488	35	33.7	464	2	C84428	probable ribophori
416	35	33.7	228	2	F72520	hypothetical prote	489	35	33.7	467	2	E69601	ATP-dependent Clp
417	35	33.7	228	2	H69025	hypothetical prote	490	35	33.7	469	1	AJZROL	glutamate-ammonia
418	35	33.7	232	2	S77110	hypothetical prote	491	35	33.7	471	1	JC2310	dihydropyrimidinas
419	35	33.7	237	2	D75359	DNA polymerase-rel	492	35	33.7	479	2	T29720	hypothetical prote
420	35	33.7	243	2	G72211	conserved hypothet	493	35	33.7	491	2	A36036	cytochrome P450 2F
421	35	33.7	248	2	C82878	integrase-recombin	494	35	33.7	494	2	B69363	hydantoin utilizat
422	35	33.7	251	2	C91263	phosphonate metabo	495	35	33.7	496	2	AH3254	acetyl-CoA:acetoac
423	35	33.7	252	2	H86103	phosphonate metabo	496	35	33.7	512	2	A98352	hydantoinase homol
424	35	33.7	252	2	H35719	phnP protein - Esc	497	35	33.7	512	2	AE2930	hydantoinase A [im
425	35	33.7	252	2	C84315	hypothetical prote	498	35	33.7	513	2	F83900	long-chain acyl-Co
426	35	33.7	254	2	H70467	HMP-P kinase - Aqu	499	35	33.7	517	2	D84421	probable amino aci
427	35	33.7	255	2	CRB02	carbonate dehydrat	500	35	33.7	528	2	I48253	beta-N-acetylhexos
428	35	33.7	259	1	E96937	chemotaxis motilit	501	35	33.7	529	1	AQHUBA	beta-N-acetylhexos
429	35	33.7	262	2	B86925	conserved hypothet	502	35	33.7	548	2	D69187	probable acid-CoA
430	35	33.7	270	2	E70669	probable glycosylt	503	35	33.7	552	2	A6756	hypothetical prote
431	35	33.7	270	2	S33440	T-cell receptor ga	504	35	33.7	555	2	E87003	probable DNA methy
432	35	33.7	274	2	G82671	protoporphyrinogen	505	35	33.7	569	2	D82824	regulator of patho
433	35	33.7	275	2	C42646	dihydropterolate sy	506	35	33.7	570	2	S09812	hypothetical prote
434	35	33.7	279	1	SYECOG	dihydropterolate sy	507	35	33.7	583	2	T39112	probable amidase -
435	35	33.7	279	2	S07654	dihydropterolate sy	508	35	33.7	608	2	I3269	prolactin receptor
436	35	33.7	279	2	T45123	dihydropterolate sy	509	35	33.7	610	2	A34631	lactogen receptor
437	35	33.7	279	2	JC5848	protein synthesis	510	35	33.7	610	2	A36116	prolactin receptor
438	35	33.7	280	2	A95936	probable alcohol s	511	35	33.7	615	2	D96499	probable UDP-gluc
439	35	33.7	283	2	S10928	dihydropterolate sy	512	35	33.7	619	2	E83635	hypothetical prote
440	35	33.7	292	2	T77525	hypothetical prote	513	35	33.7	619	2	H84415	hypothetical prote
441	35	33.7	292	2	I77525	prolactin receptor	514	35	33.7	620	2	T10423	estrogen receptor
442	35	33.7	302	2	AH0550	hypothetical ROK-f	515	35	33.7	633	2	B70946	NADH2 dehydrogenas
443	35	33.7	303	2	I77524	site-specific DNA-	516	35	33.7	650	2	S22835	alpha-agglutinin -
444	35	33.7	304	2	C64109	hypothetical prote	517	35	33.7	657	2	T34037	heat shock 70K pro
445	35	33.7	305	2	A83340	fructokinase (EC 2	518	35	33.7	659	2	S10228	paraspinal crystal
446	35	33.7	307	1	JQ0782	stage III sporulat	519	35	33.7	661	2	I56258	Rp105 - mouse
447	35	33.7	307	2	S16622	prolactin receptor	520	35	33.7	666	2	T10567	probable serine/th
448	35	33.7	310	2	A29884	hypothetical prote	521	35	33.7	666	2	T05432	hypothetical prote
449	35	33.7	315	2	AD3127	probable transcrip	522	35	33.7	677	2	T08943	hypothetical prote
450	35	33.7	315	2	E83369	probable transcrip	523	35	33.7	681	1	H82059	2',3'-cyclic-nucle
451	35	33.7	317	2	B83158	probable transcrip	524	35	33.7	682	2	T12968	hypothetical prote
452	35	33.7	318	2	S72255	ribose-phosphate d	525	35	33.7	683	1	S69780	outer membrane pro
453	35	33.7	318	2	F87529	sugar isomerase, K	526	35	33.7	690	2	H69268	copper-transportin
454	35	33.7	319	2	E97792	kpsF protein [limp	527	35	33.7	720	2	T15756	hypothetical prote
455	35	33.7	319	2	E98160	hypothetical prote	528	35	33.7	735	2	S46830	urea transport pro
456	35	33.7	322	2	B87419	3-oxoacyl-(acyl-ca	529	35	33.7	771	2	H72410	hypothetical prote
457	35	33.7	329	2	T10683	hypothetical prote	530	35	33.7	777	2	F82560	hypothetical prote
458	35	33.7	332	2	AH1562	molybdopterin bios	531	35	33.7	784	2	A86676	hypothetical prote
459	35	33.7	332	2	AD3340	cobC protein [limp	532	35	33.7	784	2	T09485	cold-induced prote
460	35	33.7	336	2	AT0183	galactoside permea	533	35	33.7	832	2	T23693	carbon starvation
461	35	33.7	338	2	E97086	anaerobic sulfite	534	35	33.7	853	2	S59315	hypothetical prote
462	35	33.7	339	2	T28019	hypothetical prote	535	35	33.7	855	2	C82983	hypothetical prote
463	35	33.7	342	2	A46396	ets-related protei	536	35	33.7	879	2	T49796	probable sepB prot
464	35	33.7	346	1	A95361	probable transcrip	537	35	33.7	902	2	T41051	beta transducin -
465	35	33.7	355	2	JQ1146	C-5 sterol desatur	538	35	33.7	935	2	E96806	hypothetical prote
466	35	33.7	373	1	H69281	conserved hypothet	539	35	33.7	973	2	A97522	ribonuclease E, RN
467	35	33.7	374	2	T27420	hypothetical prote	540	35	33.7	977	2	AC2741	ribonuclease E [im

541	35	33.7	982	2	A97210	beta galactosidase	614	34	32.7	121	2	T42138	type II secretion
542	35	33.7	1023	2	B38932	phospholipase C (E	615	34	32.7	125	2	A70356	hypothetical prote
543	35	33.7	1060	2	E83547	proline dehydrogen	616	34	32.7	133	2	AF0391	curlin genes regul
544	35	33.7	1108	2	I59385	guanylate cyclase	617	34	32.7	134	2	AG2926	conserved hypotnet
545	35	33.7	1132	2	H82887	hypothetical prote	618	34	32.7	134	2	H98355	hypothetical prote
546	35	33.7	1162	2	T49191	hypothetical prote	619	34	32.7	141	1	HADD	hemoglobin alpha c
547	35	33.7	1163	2	JF0366	tight junction pro	620	34	32.7	141	2	T20663	hypothetical prote
548	35	33.7	1169	2	T18423	hypothetical prote	621	34	32.7	146	2	S36323	T-cell receptor de
549	35	33.7	1171	2	A80130	DNA-directed DNA p	622	34	32.7	147	2	D75132	probable hydrogena
550	35	33.7	1172	2	T49330	cytokinesis inhibi	623	34	32.7	158	2	G75542	probable RNA methy
551	35	33.7	1176	1	A40447	phospholipase C (E	624	34	32.7	163	2	H69990	sugar transport pr
552	35	33.7	1179	2	T04488	DNA topoisomerase	625	34	32.7	167	2	AE2342	hypothetical prote
553	35	33.7	1186	2	AD1300	Smc protein essent	626	34	32.7	167	2	E71943	hypothetical prote
554	35	33.7	1186	2	AD1672	Smc protein essent	627	34	32.7	170	2	T42413	FMRFamide-like pep
555	35	33.7	1205	2	B95845	conserved hypotnet	628	34	32.7	175	2	AF1512	hypothetical prote
556	35	33.7	1208	2	AI3584	N-methylhydantoina	629	34	32.7	179	2	A95245	hypothetical prote
557	35	33.7	1217	2	T00270	hypothetical prote	630	34	32.7	183	2	G96018	probable methylate
558	35	33.7	1275	2	T18556	O-antigen biosynth	631	34	32.7	199	2	A82360	conserved hypotnet
559	35	33.7	1321	2	A60165	sodium channel pro	632	34	32.7	207	2	G95238	conserved hypotnet
560	35	33.7	1341	2	S50366	probable membrane	633	34	32.7	207	2	H98102	hypothetical prote
561	35	33.7	1386	2	T49316	profilaggrin relat	634	34	32.7	208	2	T20554	hypothetical prote
562	35	33.7	1413	2	D84481	probable retroelem	635	34	32.7	213	2	I39900	conserved hypotnet
563	35	33.7	1498	2	AG1439	B. subtilis Yuka p	636	34	32.7	214	2	C82255	adenylate kinase V
564	35	33.7	1832	2	T31113	mucin-like glycopr	637	34	32.7	216	1	BVECKB	alkB protein - Esc
565	35	33.7	2126	2	H70621	probable polyketid	638	34	32.7	216	2	E91016	alkylated DNA repa
566	35	33.7	2179	1	GNNYH4	genome polyprotein	639	34	32.7	216	2	G85860	DNA repair system
567	35	33.7	2262	2	T30890	calcium channel al	640	34	32.7	219	2	AG0872	probable membrane
568	35	33.7	2337	2	T40577	probable phosphati	641	34	32.7	219	2	C65074	probable oxidoredu
569	35	33.7	2493	2	S45734	probable membrane	642	34	32.7	219	2	C65074	probable oxidoredu
570	35	33.7	3169	2	T00296	toxin B - Escheric	643	34	32.7	219	2	S59967	hemolysin III - Ba
571	34.5	33.2	142	2	S36310	T-cell receptor de	644	34	32.7	219	2	G85945	probable oxidoredu
572	34.5	33.2	238	1	TRWV5Y	trypsin-like prote	645	34	32.7	222	2	B54898	STX protein - huma
573	34.5	33.2	255	2	AE0537	hypothetical prote	646	34	32.7	225	2	A35295	glutathione transf
574	34.5	33.2	32	2	T24658	hypothetical prote	647	34	32.7	228	2	B86836	hypothetical prote
575	34.5	33.2	290	2	A46353	ORF1 protein - coc	648	34	32.7	231	2	JQ1247	hypothetical 25.7K
576	34.5	33.2	322	1	A70189	hypothetical prote	649	34	32.7	235	2	A97996	degenerate transpo
577	34.5	33.2	326	2	AF2024	hypothetical prote	650	34	32.7	241	2	AG2915	conserved hypotnet
578	34.5	33.2	334	2	AG2190	hypothetical prote	651	34	32.7	241	2	B97690	hypothetical prote
579	34.5	33.2	338	2	T23520	hypothetical prote	652	34	32.7	243	2	C95908	hypothetical prote
580	34.5	33.2	394	2	G86657	ABC transporter At	653	34	32.7	247	2	C87423	cytochrome c oxida
581	34.5	33.2	412	2	T40295	fructosyl amine -	654	34	32.7	253	2	H69046	hypothetical prote
582	34.5	33.2	450	2	G33740	hypothetical prote	655	34	32.7	256	1	A42768	homeotic protein g
583	34.5	33.2	466	2	AI1213	TN916 ORF21 homolo	656	34	32.7	256	2	AG2873	dehydrogenase Atu2
584	34.5	33.2	511	2	AB0858	hypothetical prote	657	34	32.7	256	2	H97649	probable gluconate
585	34.5	33.2	538	2	T06683	aldehyde dehydroge	658	34	32.7	256	2	AB1098	conserved hypotnet
586	34.5	33.2	570	2	T02676	hypothetical prote	659	34	32.7	257	2	AB1098	conserved hypotnet
587	34.5	33.2	600	1	D65000	NADH2 dehydrogenas	660	34	32.7	257	2	AI1460	conserved hypotnet
588	34.5	33.2	600	2	AD0797	NADH2 dehydrogenas	661	34	32.7	259	1	CR8B2	carbonate dehydrat
589	34.5	33.2	600	2	C85869	NADH dehydrogenase	662	34	32.7	260	1	CRU2	carbonate dehydrat
590	34.5	33.2	600	2	B91025	NADH dehydrogenase	663	34	32.7	261	2	T22623	hypothetical prote
591	34.5	33.2	603	2	T02677	hypothetical prote	664	34	32.7	263	2	B82731	UDP-N-acetylglucos
592	34.5	33.2	605	1	A35459	glucose oxidase (E	665	34	32.7	266	2	B89886	glutamate racemase
593	34.5	33.2	611	2	T28171	hypothetical prote	666	34	32.7	268	2	G02133	holocytochrome-c s
594	34.5	33.2	612	2	A44857	acetylactate synth	667	34	32.7	268	2	T27242	hypothetical prote
595	34.5	33.2	635	1	A29358	cerevisin (EC 3.4.	668	34	32.7	270	2	T09514	5'-AMP-activated p
596	34.5	33.2	693	2	AF2357	hypothetical prote	669	34	32.7	276	2	D70191	hypothetical prote
597	34.5	33.2	705	1	TFCHE	ovotransferrin pre	670	34	32.7	280	2	E75216	hypothetical prote
598	34.5	33.2	856	2	A64699	hypothetical prote	671	34	32.7	280	2	C71453	hypothetical prote
599	34.5	33.2	967	2	C70831	probable mmp14 pro	672	34	32.7	282	2	E72415	zinc ABC transport
600	34.5	33.2	968	2	C70746	probable mmp12 pro	673	34	32.7	288	2	S03603	uncoupling protein
601	34.5	33.2	1744	1	C4HU	complement C4a pre	674	34	32.7	291	2	A90084	hypothetical prote
602	34	32.7	55	2	AE2103	hypothetical prote	675	34	32.7	294	2	C72689	probable diptine
603	34	32.7	61	2	B4091	hypothetical prote	676	34	32.7	299	2	AC2998	glutamyl-trna synt
604	34	32.7	65	2	B38601	Ig kappa chain V r	677	34	32.7	299	2	F98285	glutamyl-trna synt
605	34	32.7	85	2	S25837	hypothetical prote	678	34	32.7	309	2	T12089	HA-exporting ANPas
606	34	32.7	93	2	A38601	Ig kappa chain V r	679	34	32.7	309	2	T70408	hypothetical prote
607	34	32.7	98	2	S44785	C30C11.3 protein -	680	34	32.7	310	2	T09562	beta-carotene hydr
608	34	32.7	104	2	K1HURY	Ig kappa chain V-I	681	34	32.7	311	2	T15997	hypothetical prote
609	34	32.7	113	2	E44151	Ig lambda chain V	682	34	32.7	312	2	T02406	hypothetical prote
610	34	32.7	113	2	T00196	hypothetical prote	683	34	32.7	313	2	H70777	probable cobd - My
611	34	32.7	116	2	T42412	FMRFamide-like pep	684	34	32.7	325	1	A48561	inner capsid prote
612	34	32.7	117	2	AH0623	probable secreted	685	34	32.7	335	2	G75040	probable aminopept
613	34	32.7	121	2	S26798	Ig heavy chain V r	686	34	32.7	335	2	B98142	hypothetical prote

687	34	32.7	340	2	G75531	cytochrome c famil	760	34	32.7	484	1	A35282	alpha-amylase (EC
688	34	32.7	343	2	H96984	endoglucanase, ami	761	34	32.7	486	1	S63251	homoserine O-acetyl
689	34	32.7	345	2	T16427	hypothetical prote	762	34	32.7	489	2	T09151	glutathione-disulf
690	34	32.7	347	2	B87274	hypothetical prote	763	34	32.7	499	2	B83493	conserved hypothet
691	34	32.7	356	2	A13145	GDP-mannose 4,6-de	764	34	32.7	504	2	A83286	acetyl-CoA C-acety
692	34	32.7	357	1	E64649	probable X-Pro dip	765	34	32.7	506	2	S63181	hypothetical prote
693	34	32.7	357	2	A71938	probable proline p	766	34	32.7	508	2	AG2958	hypothetical prote
694	34	32.7	357	2	B98133	hypothetical prote	767	34	32.7	509	1	O4FTLO	laurate omega-hydr
695	34	32.7	357	2	AG3154	conserved hypothet	768	34	32.7	513	2	H83375	probable glycogen
696	34	32.7	359	2	F71136	hypothetical prote	769	34	32.7	517	2	G98324	hypothetical prote
697	34	32.7	363	2	AB2065	threonine synthase	770	34	32.7	518	2	A12929	hydantoinase [limpo
698	34	32.7	373	2	T48381	hypothetical prote	771	34	32.7	518	2	E98352	hydantoinase [limpo
699	34	32.7	375	2	T02215	ferredoxin-NADP re	772	34	32.7	524	2	T09937	cytokinin oxidase
700	34	32.7	375	2	I39169	sialyltransferase	773	34	32.7	529	1	H70736	IMP dehydrogenase
701	34	32.7	375	2	I48686	N-glycan alpha 2,8	774	34	32.7	530	2	T43089	transfer complex p
702	34	32.7	375	2	A46727	sialyltransferase	775	34	32.7	533	2	D83885	hydantoinase BH188
703	34	32.7	378	2	T02977	ferredoxin-NADP re	776	34	32.7	542	2	T02379	hypothetical prote
704	34	32.7	380	2	A83834	transposase (02) B	777	34	32.7	544	2	C82900	probable ABC subst
705	34	32.7	380	2	AB1709	N-acetylglucosamin	778	34	32.7	550	1	VGBE18	glycoprotein E - h
706	34	32.7	382	2	B86430	hypothetical prote	779	34	32.7	555	2	D95139	DNA repair protein
707	34	32.7	383	1	VGBEKG	glycoprotein precu	780	34	32.7	555	2	C98007	DNA repair and gen
708	34	32.7	383	2	H95970	probable sugar upt	781	34	32.7	573	2	S68591	methanol dehydroge
709	34	32.7	391	2	T27918	hypothetical prote	782	34	32.7	578	2	T11659	hypothetical prote
710	34	32.7	393	2	T23602	hypothetical prote	783	34	32.7	580	2	S63391	probable membrane
711	34	32.7	394	2	G72212	exodeoxyribonuclea	784	34	32.7	581	2	C39768	cholinesterase (EC
712	34	32.7	395	2	B47071	conserved hypothet	785	34	32.7	600	2	A49230	cholinesterase (EC
713	34	32.7	396	2	AF0176	nitrogenase cofact	786	34	32.7	602	1	ACHU	dnak-type molecula
714	34	32.7	403	2	A28451	beta-galactoside a	787	34	32.7	602	2	A53583	geranylgeranyl-dip
715	34	32.7	404	2	B83769	transposase (02) B	788	34	32.7	608	2	F83397	probable peptidase
716	34	32.7	404	2	G83797	transposase (02) B	789	34	32.7	611	2	D70928	hypothetical prote
717	34	32.7	404	2	B83817	transposase (02) B	790	34	32.7	615	2	G82658	periplasmic glucan
718	34	32.7	404	2	F83860	transposase (02) B	791	34	32.7	627	2	T32958	hypothetical prote
719	34	32.7	404	2	E83862	transposase (02) B	792	34	32.7	637	2	JQ2039	polyprotein - pars
720	34	32.7	404	2	C83895	transposase (02) B	793	34	32.7	639	2	B86587	DNA helicase [limpo
721	34	32.7	404	2	D83925	transposase (02) B	794	34	32.7	639	2	G72038	ATP-dependent heli
722	34	32.7	404	2	B83947	transposase (02) B	795	34	32.7	641	1	UNBPAL	DNA-packaging prot
723	34	32.7	404	2	H83958	transposase (02) B	796	34	32.7	641	2	JN0668	dnak-type molecula
724	34	32.7	404	2	E83959	transposase (02) B	797	34	32.7	641	2	H85689	terminase large su
725	34	32.7	404	2	A83980	transposase (02) B	798	34	32.7	641	2	F90832	terminase large su
726	34	32.7	404	2	C83981	transposase (02) B	799	34	32.7	641	2	C90901	probable terminase
727	34	32.7	404	2	B84015	transposase (02) B	800	34	32.7	650	2	S75072	probable phytoene
728	34	32.7	404	2	B84022	transposase (02) B	801	34	32.7	650	2	G83465	conserved hypothet
729	34	32.7	404	2	G84055	transposase (02) B	802	34	32.7	654	2	H86717	hypothetical prote
730	34	32.7	404	2	A84076	transposase (02) B	803	34	32.7	656	1	A34890	histidine decarbox
731	34	32.7	404	2	C84088	transposase (02) B	804	34	32.7	658	2	S39653	xcpQ protein - Pse
732	34	32.7	404	2	E84095	transposase (02) B	805	34	32.7	662	1	A49882	histidine decarbox
733	34	32.7	405	2	T27971	hypothetical prote	806	34	32.7	662	1	S12889	histidine decarbox
734	34	32.7	406	2	S68866	qsopA protein - Co	807	34	32.7	662	2	G95929	probable methylcro
735	34	32.7	420	2	AI3614	hypothetical prote	808	34	32.7	662	2	E97973	hypothetical prote
736	34	32.7	422	2	C70370	dihydroorotase - A	809	34	32.7	685	2	T09159	RNA helicase prh75
737	34	32.7	425	2	C87563	homogentisate 1,2-	810	34	32.7	698	2	B96958	dnak protein (heat
738	34	32.7	427	2	D81784	probable integral	811	34	32.7	704	2	H95890	probable hydantoin
739	34	32.7	427	2	G81208	AmpG-related prote	812	34	32.7	704	2	AH1958	hypothetical prote
740	34	32.7	428	2	T05521	alpha-amylase (EC	813	34	32.7	741	2	T12762	NADH2 dehydrogenas
741	34	32.7	430	2	F87711	heat shock protein	814	34	32.7	751	1	SYBYMT	methionine-tRNA li
742	34	32.7	434	2	AB3508	heat shock protein	815	34	32.7	758	1	SUBCCA	ATP-dependent Clp
743	34	32.7	437	2	T37469	homogentisate 1,2-	816	34	32.7	758	2	AI0609	ATP-binding compon
744	34	32.7	449	2	C97757	hypothetical prote	817	34	32.7	758	2	H90749	ATP-binding compon
745	34	32.7	449	2	T19626	hypothetical prote	818	34	32.7	758	2	D85600	Shaw type potassi
746	34	32.7	450	2	B70318	chaperone HslU - A	819	34	32.7	769	2	I56546	Imbition protein
747	34	32.7	450	2	B97754	heat shock protein	820	34	32.7	773	2	T46188	glycoprotein X pre
748	34	32.7	450	2	B71688	heat shock protein	821	34	32.7	797	1	VBEX1	GTP diphosphokinas
749	34	32.7	458	2	S44741	Co2C2_3 protein -	822	34	32.7	847	2	S70687	hypothetical prote
750	34	32.7	461	2	AE0311	NADH2 dehydrogenas	823	34	32.7	862	2	F75116	membrane glycoprot
751	34	32.7	465	2	D84427	hypothetical prote	824	34	32.7	866	2	T45462	membrane glycoprot
752	34	32.7	466	2	T28216	hypothetical prote	825	34	32.7	867	2	T45463	hypothetical prote
753	34	32.7	469	2	AC2794	glutamine syntheta	826	34	32.7	867	2	T14777	replicative DNA he
754	34	32.7	469	2	B97573	glutamine syntheta	827	34	32.7	879	2	A11878	hypothetical prote
755	34	32.7	469	2	B75039	hypothetical prote	828	34	32.7	882	2	A39030	androgen-binding p
756	34	32.7	471	2	AI0362	glutamate-tRNA lig	829	34	32.7	895	2	B64238	isoleucine-tRNA li
757	34	32.7	473	2	AF2433	aldehyde dehydroge	830	34	32.7	896	1	A35782	cytokine receptor
758	34	32.7	473	2	T04482	ribophorin I homol	831	34	32.7	905	2	G84582	hypothetical prote
759	34	32.7	483	2	T21327	hypothetical prote	832	34	32.7	914	2	E83901	hypothetical prote

833	34	32.7	934	2	B83789	hypothetical prote	906	33.5	32.2	428	2	B82248	long-chain fatty a
834	34	32.7	935	2	A40694	cadherin-associate	907	33.5	32.2	431	2	A56822	synapconemal compl
835	34	32.7	948	2	T47322	plasma membrane H+	908	33.5	32.2	439	2	T02453	hypothetical prote
836	34	32.7	949	2	G84486	probable plasma me	909	33.5	32.2	486	2	T18903	hypothetical prote
837	34	32.7	953	2	S55156	probable membrane	910	33.5	32.2	551	2	D64412	hypothetical prote
838	34	32.7	966	2	F84582	hypothetical prote	911	33.5	32.2	556	2	T39109	probable guanosine
839	34	32.7	981	2	T46330	hypothetical prote	912	33.5	32.2	593	2	D83316	NADH dehydrogenase
840	34	32.7	990	2	T05197	hypothetical prote	913	33.5	32.2	599	2	B72368	conserved hypotnet
841	34	32.7	1014	2	T13476	hypothetical prote	914	33.5	32.2	602	2	T47794	hypothetical prote
842	34	32.7	1017	2	T31354	hypothetical prote	915	33.5	32.2	632	2	T48316	hypothetical prote
843	34	32.7	1034	2	T30574	probable potassium	916	33.5	32.2	686	2	AI0020	probable membrane
844	34	32.7	1034	2	JC5598	beta-galactosidase	917	33.5	32.2	794	2	B95942	probable aldehyde
845	34	32.7	1042	2	A57534	mucin - rat	918	33.5	32.2	811	2	T19974	hypothetical prote
846	34	32.7	1063	2	D83789	mucin 5AC (clone L	919	33.5	32.2	816	2	S19139	hypothetical prote
847	34	32.7	1106	2	S38783	hypothetical prote	920	33.5	32.2	859	2	F69159	sucrose synthase (
848	34	32.7	1123	2	A72311	integrin alpha cha	921	33.5	32.2	939	2	T32521	protoporphyrin IX
849	34	32.7	1131	2	F88570	conserved hypotnet	922	33.5	32.2	1904	2	T13256	hypothetical prote
850	34	32.7	1152	2	AE1852	protein K03H1.2 [i	923	33.5	32.2	2344	2	T41590	tail-host specific
851	34	32.7	1165	2	S45879	hypothetical prote	924	33.5	32.2	2767	1	UIHU	probable sensor-li
852	34	32.7	1442	2	T42607	chitin synthase [E	925	33	31.7	49	2	B82007	thyroglobulin prec
853	34	32.7	1451	2	S41025	transcription acti	926	33	31.7	57	2	D81226	hypothetical prote
854	34	32.7	1587	2	AB2012	hypothetical prote	927	33	31.7	61	2	H95874	hypothetical prote
855	34	32.7	1621	2	A92255	hypothetical prote	928	33	31.7	64	2	S09280	hypothetical prote
856	34	32.7	1708	1	A43100	hypothetical prote	929	33	31.7	68	2	H98098	ferredoxin 2[4Fe-4
857	34	32.7	1873	2	A55645	ataxia telangiecta	930	33	31.7	74	2	T07559	hypothetical prote
858	34	32.7	1879	2	T19481	calcium channel, v	931	33	31.7	76	2	T28248	hypothetical prote
859	34	32.7	1891	2	T43262	hypothetical prote	932	33	31.7	81	2	G85844	ORF MSV087 probabl
860	34	32.7	2476	2	T34022	calcium channel al	933	33	31.7	88	1	EDB51	unknown protein en
861	34	32.7	4427	2	PN0637	zonadhesin - pig	934	33	31.7	90	2	A69323	immediate-early-5
862	34	32.7	5255	2	T31677	polyketide synthas	935	33	31.7	99	2	S20342	hypothetical prote
863	33.5	32.2	62	2	T36967	bacitracin synthet	936	33	31.7	108	2	S64602	calcium-binding pr
864	33.5	32.2	62	2	A00880	gene MHC DQ-alpha	937	33	31.7	110	2	T10451	hypothetical prote
865	33.5	32.2	63	2	I36900	hypothetical prote	938	33	31.7	120	2	S36306	hypothetical prote
866	33.5	32.2	63	2	I61798	gene MHC DQ-alpha	939	33	31.7	120	2	S57405	T-cell receptor de
867	33.5	32.2	63	2	I61801	gene MHC DQ-alpha	940	33	31.7	121	2	E64391	response regulator
868	33.5	32.2	63	2	I61787	gene MHC DQ-alpha	941	33	31.7	121	2	AI1646	hypothetical prote
869	33.5	32.2	63	2	I61789	gene MHC DQ-alpha	942	33	31.7	124	2	B86802	hypothetical prote
870	33.5	32.2	63	2	H34513	gene MHC DQ-alpha	943	33	31.7	133	2	S15472	prophage p13 prote
871	33.5	32.2	63	2	G34513	MHC class II histo	944	33	31.7	133	2	D72110	hypothetical prote
872	33.5	32.2	63	2	B34513	MHC class II histo	945	33	31.7	133	2	E86513	hypothetical prote
873	33.5	32.2	63	2	C34514	MHC class II histo	946	33	31.7	138	2	JC4597	hypothetical prote
874	33.5	32.2	87	2	C30575	MHC class II histo	947	33	31.7	141	1	HATG1	signal transductio
875	33.5	32.2	87	2	B27628	MHC class II histo	948	33	31.7	141	1	HAWHK	hemoglobin alpha-i
876	33.5	32.2	142	2	S17462	glutathione transf	949	33	31.7	141	2	A25728	hemoglobin alpha c
877	33.5	32.2	158	2	G95925	glutathione transf	950	33	31.7	142	2	S04071	hemoglobin alpha c
878	33.5	32.2	183	2	A99262	probable acetyltra	951	33	31.7	145	2	H89587	hemoglobin alpha c
879	33.5	32.2	205	2	F36138	hypothetical prote	952	33	31.7	145	2	A41897	protein R09F10.5 [
880	33.5	32.2	205	2	A85604	urase accessory p	953	33	31.7	149	2	A26042	cellulase homolog
881	33.5	32.2	205	2	G90794	probable urease ac	954	33	31.7	153	2	E86872	globin 1 - sea lam
882	33.5	32.2	212	2	I46176	urase accessory p	955	33	31.7	157	2	G75560	galactoside O-acet
883	33.5	32.2	218	1	XURTG4	thyroglobulin prec	956	33	31.7	157	2	B69817	conserved hypotnet
884	33.5	32.2	218	2	S01719	glutathione transf	957	33	31.7	158	1	A42082	hypothetical prote
885	33.5	32.2	218	2	A23732	glutathione transf	958	33	31.7	159	2	T03958	protein-tyrosine-p
886	33.5	32.2	218	2	B34159	glutathione transf	959	33	31.7	160	2	E87384	heat shock protein
887	33.5	32.2	220	2	JQ1986	glutathione transf	960	33	31.7	164	2	B73311	transcription regu
888	33.5	32.2	234	2	S32742	avrFmaA1 protein -	961	33	31.7	169	2	B81325	conserved signal-tr
889	33.5	32.2	243	2	A36124	genome polypeptin	962	33	31.7	171	2	A96901	conserved hypotnet
890	33.5	32.2	243	2	I54290	thyroglobulin 2 pr	963	33	31.7	172	2	G82987	uncharacterized co
891	33.5	32.2	255	1	HLHUD1	cell surface glyco	964	33	31.7	172	2	T06832	secreted protein H
892	33.5	32.2	266	2	A72401	MHC class II histo	965	33	31.7	176	2	S23343	ycf37 protein - Cy
893	33.5	32.2	267	2	G70455	ABC transporter, A	966	33	31.7	181	2	E55213	hypothetical prote
894	33.5	32.2	288	2	D87706	hypothetical prote	967	33	31.7	183	2	S78545	rbbd protein - Shi
895	33.5	32.2	288	2	E90172	diaminopelate ep	968	33	31.7	183	2	S69432	trpP-4-dehydroxam
896	33.5	32.2	301	2	D81745	hypothetical prote	969	33	31.7	184	2	AI3539	hypothetical prote
897	33.5	32.2	305	2	S38817	probable phosphati	970	33	31.7	185	2	B97443	hypothetical prote
898	33.5	32.2	320	2	A69789	FMRFamide-related	971	33	31.7	185	2	A26661	hypothetical prote
899	33.5	32.2	387	2	AI1050	fructokinase homol	972	33	31.7	189	2	D69389	conserved hypotnet
900	33.5	32.2	411	2	T15705	conserved hypotnet	973	33	31.7	189	2	T17571	DNA-directed RNA p
901	33.5	32.2	415	2	D71936	hypothetical prote	974	33	31.7	190	2	H64202	hypothetical prote
902	33.5	32.2	417	2	JH0660	hypothetical prote	975	33	31.7	192	2	S36199	translation elonga
903	33.5	32.2	422	2	AG2999	amine dehydrogenas	976	33	31.7	194	2	T44442	hypothetical prote
904	33.5	32.2	422	2	B98284	diaminopelate de	977	33	31.7	196	2	AE0077	conserved hypotnet
905	33.5	32.2	426	2	A36934	diaminopelate de	978	33	31.7	201	2	T37036	conserved hypotnet
						amine dehydrogenas							

979 33 31.7 202 2 A89911 conserved hypothet
980 33 31.7 202 2 T28357 ORF MSV196 Ali mot
981 33 31.7 202 2 F71293 hypothetrical prote
982 33 31.7 203 2 T41946 hypothetrical prote
983 33 31.7 219 2 AG2124 hypothetrical prote
984 33 31.7 220 2 AC0318 probable nicotinat
985 33 31.7 223 2 F91210 hypothetrical prote
986 33 31.7 223 2 A86057 hypothetrical prote
987 33 31.7 224 2 D84345 phosphoribosylform
988 33 31.7 230 1 WNWV26 p26 protein [simil
989 33 31.7 233 2 A11152 hypothetrical prote
990 33 31.7 233 2 AB1512 hypothetrical prote
991 33 31.7 233 2 F82379 probable chemotaxi
992 33 31.7 239 2 H90218 SSU ribosomal prot
993 33 31.7 240 2 S71458 ABC-type heme tran
994 33 31.7 240 2 S71456 ABC-type heme tran
995 33 31.7 240 2 S62088 probable transport
996 33 31.7 240 2 A43912 myogenin - Japanes
997 33 31.7 242 2 A82754 lipoprotein XF0855
998 33 31.7 243 2 A95162 alpha-acetolactate
999 33 31.7 243 2 H98027 acetolactate decar
1000 33 31.7 244 2 F84969 flagellar basal-bo

ALIGNMENTS

RESULT 1
JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Pa
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family m
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI0000003888A; GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and thera
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:Keywords: heparin binding; membrane bound

Query Match 63.5%; Score 66; DB 2; Length 480;
Best Local Similarity 75.0%; Pred. No. 0.0074;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVLRFGGTKTDFLIIF 16
Db 110 PAVLRFGGKRTDFLQF 125
||:|||||:|||||
RESULT 2
D70394
mannosyltransferase A - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: D70394
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: D70394
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-588 <AQF>
A:Cross-references: UNIPROT:O67183; UNIPARC:UPI0000005651A; GB:AE000723; NID:g2983569; PJ

A:Experimental source: strain VFS
C:Genetics:
A:Gene: mtfa

Query Match 47.1%; Score 49; DB 2; Length 588;
Best Local Similarity 60.0%; Pred. No. 5.5;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YLRFGGTKTDFLIIF 17
Db 307 YLKFKTKTFLVLYD 321
||:|||||:|||||

RESULT 3

CRH07

carbonate dehydratase (EC 4.2.1.1) VII - human

N:Alternate names: carbonic anhydrase VII

C:Species: Homo sapiens (man)

C:Date: 10-Feb-1995 #sequence_revision 05-May-1995 #text_change 09-Jul-2004

C:Accession: A55272

R:Montgomery, J.C.; Venta, P.J.; Eddy, R.L.; Fukushima, Y.S.; Shows, T.B.; Tashian, R.E.

Genomics 11, 835-848, 1991

A:Title: Characterization of the human gene for a newly discovered carbonic anhydrase, C

A:Reference number: A55272; MUID:92147127; PMID:1783392

A:Accession: A55272

A:Molecule type: DNA

A:Residues: 1-264 <MON>

A:Cross-references: UNIPROT:P43166; UNIPARC:UPI000000D814; GB:M76423; NID:gl79964; PIDN:/

A:Note: sequence extracted from NCBI backbone (NCBIN:80199, NCBIN:80201, NCBIN:80205, NCI

C:Genetics:

A:Gene: GDB:CA7

A:Cross-references: GDB:119741; OMIM:114770

A:Map position: 16q22.1-16q22.1

A:Introns: 14/1; 80/1; 119/3; 151/3; 172/3; 224/3

C:Function:

A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide

A:Note: This form is expressed in salivary gland and other tissues

C:Superfamily: carbonate dehydratase; carbonic anhydrase homology

C:Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc

E:2-264/Product: carbonate dehydratase VII #status predicted <MAT>

E:7-262/Domain: carbonic anhydrase homology <CAH>

E:796,98,121/Binding site: zinc (His) #status predicted

Query Match 46.2%; Score 48; DB 1; Length 264;

Best Local Similarity 56.2%; Pred. No. 3.5;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLIIFDPK 19
Db 169 VRFGTKAQFSCFNPK 184
||:|||||:|||||

RESULT 4

C86516

hypothetical protein CPJ0203 [imported] - Chlamydomophila pneumoniae (strain J138)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: C86516

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; I

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: C86516

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <STO>

A:Cross-references: UNIPROT:Q9Z8Y1; UNIPARC:UPI00000470E1; GB:BA000008; NID:g8978576; PII

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0203

Query Match 46.2%; Score 48; DB 2; Length 265;

Best Local Similarity 47.1%; Pred. No. 3.5;

Matches	8;	Conservative	3;	Mismatches	6;	Indels	0;	Gaps	0;
Qy	2	AYLRFGGTKTDFLI	FDP 18						
	:	:	:						
Db	24	SYFFFGGTRTQIL	VI	TP 40					
RESULT 5									
D72105									
KDO-transferrase 2, probable frameshift CP0564 [imported] - Chlamydia pneumoniae (strain 12B)									
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae									
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004									
C:Accession: D72105; D81563									
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;									
Nature Genet. 21, 385-389, 1999									
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.									
A:Reference number: A72000; MUID:99206606; PMID:10192388									
A:Accession: D72105									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-265 <ARN>									
A:Cross-references: UNIPROT:Q9Z8Y1; UNIPARC:UPI00000470E1; GB:AE001607; GB:AE001363; NID:									
A:Experimental source: strain CWL029									
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,									
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,									
Nucleic Acids Res. 28, 1397-1406, 2000									
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.									
A:Reference number: A81500; MUID:20150255; PMID:10684935									
A:Accession: D81563									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-265 <REA>									
A:Cross-references: UNIPARC:UPI00000470E1; GB:AE002215; GB:AE002161; NID:g7189472; PIDN:									
A:Experimental source: strain AR39, HL cells									
C:Genetics:									
A:Gene: CPN0203; CP0564									
Query Match 46.2%; Score 48; DB 2; Length 265;									
Best Local Similarity 47.1%; Pred. No. 3.5;									
Matches	8;	Conservative	3;	Mismatches	6;	Indels	0;	Gaps	0;
Qy	2	AYLRFGGTKTDFLI	FDP 18						
	:	:	:						
Db	24	SYFFFGGTRTQIL	VI	TP 40					
RESULT 6									
E72351									
Hypothetical protein TW0646 - Thermotoga maritima (strain MSB8)									
C:Species: Thermotoga maritima									
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004									
C:Accession: E72351									
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,									
Garrett, M.N.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;									
C.M.									
Nature 399, 323-329, 1999									
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq									
A:Reference number: A72200; MUID:99287316; PMID:10360571									
A:Accession: E72351									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-176 <ARN>									
A:Cross-references: UNIPROT:Q9WZB4; UNIPARC:UPI000000C13C4; GB:AE001738; GB:AE000512; NID:									
A:Experimental source: strain MSB8									
C:Genetics:									
A:Gene: TW0646									
C:Superfamily: Thermotoga maritima hypothetical protein TW0646									
Query Match 45.2%; Score 47; DB 2; Length 176;									
Best Local Similarity 52.9%; Pred. No. 3.3;									
Matches	9;	Conservative	2;	Mismatches	6;	Indels	0;	Gaps	0;
Qy	1	PAYLRFGGTKTDFLI	FDP 17						

A;Accession: G64971
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Cross-references: 1-405 <BLAT>
A;Cross-references: UNIPROT:P71238; UNIPARC:UPI0000138EC5; GB:AE000295; GB:U00096; NID:9
A;Experimental source: strain K-12, substrain WGL655

Query Match 44.2%; Score 46; DB 2; Length 405;
Best Local Similarity 43.8%; Pred. No. 12;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 YLRFGGTYTDFLIFDP 18
::|||:|:|:
Db 162 FIKFGKRTTALYFEP 177

RESULT 10
S77596
cytochrome-c oxidase (EC 1.9.3.1) type cbb3 chain fixO - Paracoccus denitrificans
C;Species: Paracoccus denitrificans
C;Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 09-Jul-2004
C;Accession: S77596
R;de Gier, J.W.; Schepper, M.; Reijnders, W.N.M.; van Dyck, S.J.; Slotboom, D.J.; Warne, C.
Mol. Microbiol. 20, 1247-1260, 1996
A;Title: Structural and functional analysis of aa(3)-type and cbb(3)-type cytochrome c
A;Reference number: S77595; MUID:96405647; PMID:8809776
A;Accession: S77596
A;Molecule type: DNA
A;Residues: 1-241 <DEA>
A;Cross-references: UNIPROT:Q51680; UNIPARC:UPI00000082F1; EMBL:U34353; NID:gl002874; PDB:1
A;Experimental source: strain Pd1222
C;Genetics:
A;Gene: ccoO
C;Function:
A;Pathway: respiratory chain
C;Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C;Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 43.3%; Score 45; DB 2; Length 241;
Best Local Similarity 52.9%; Pred. No. 9.8;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGTYKTDFLIFDP 18
|||:|:|:|:
Db 221 AYLVGLTGWDFSTFEP 237

RESULT 11
B55582
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Azorhizobium caulinodans
N;Alternate names: cb-type cytochrome-c oxidase 28K chain; cytochrome b410; fixO protein
C;Species: Azorhizobium caulinodans
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: B55582; S42230
R;Mandon, K.; Kaminski, P.A.; Elmerich, C.
J. Bacteriol. 176, 2560-2568, 1994
A;Title: Functional analysis of the fixNOQP region of Azorhizobium caulinodans.
A;Reference number: A55582; MUID:94222833; PMID:8169204
A;Accession: B55582
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-246 <MAN1>
A;Cross-references: UNIPROT:Q43943; UNIPARC:UPI0000085F6A; GB:X74410; NID:g456310; PIDN:R;Mandon, K.; Kaminski, P.A.; Mougel, C.; Desnoues, N.; Dreyfus, B.; Elmerich, C.
FEMS Microbiol. Lett. 114, 185-190, 1993
A;Title: Role of the fixGHI region of Azorhizobium caulinodans in free-living and symbio
A;Reference number: S42229; MUID:94109675; PMID:8282187
A;Accession: S42230
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-246 <MAN2>
A;Cross-references: UNIPARC:UPI0000085F6A; EMBL:X74410; NID:g456310; PIDN:CAAS52430.1; PDB:1
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993

C;Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C;Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 43.3%; Score 45; DB 1; Length 246;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIFDPK 19
|||: ||: ||: ||: ||
Db 225 AYLQQLGTQVDFKLYDNK 242

RESULT 12
F84363
Brp-like homolog [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: F84363
R;NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: F84363
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-284 <STO>
A;Cross-references: UNIPROT:Q9HNE6; UNIPARC:UPI0000063A63; GB:AE004437; NID:gl0581551; P1
C;Genetics:
A;Gene: blh

Query Match 42.3%; Score 44; DB 2; Length 284;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 6 FGGTKTDFLIFDPK 19
|||: ||: ||: ||: ||
Db 118 FGGSGVDLAVFDPK 131

RESULT 13
D87403
cbbC protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 05-Oct-2004
C;Accession: D87403
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87403
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-762 <STO>
A;Cross-references: UNIPROT:Q9A8V8; UNIPARC:UPI00000C7309; GB:AE005673; NID:gl3422572; P1
C;Genetics:
A;Gene: CC1243
C;Superfamily: molybdopterin dependent formate dehydrogenase

Query Match 42.3%; Score 44; DB 2; Length 762;
Best Local Similarity 53.3%; Pred. No. 47;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 5 RFGGKTDFLIFDPK 19
|||: ||: ||: ||: ||
Db 617 RTASCKANFLVDPK 631

RESULT 14

JC6067
CCAAT-binding factor CBF1 - mouse
N:Alternate names: CBF1 protein
C:Species: Mus musculus (house mouse)
C>Date: 21-Jan-1997 #sequence_revision 21-Jan-1997 #text_change 09-Jul-2004
R:Hoepfner, M.A.; Gilbert, D.J.; Copeland, N.G.; Jenkins, N.A.; Linzer, D.I.H.; Wu, B.
Nucleic Acids Res. 24, 1091-1098, 1996
A:Title: Cloning and characterization of mouse CCAAT binding factor.
A:Reference number: S65567; MUID:96184859; PMID:8604343
A:Accession: JC6067
A:Molecule type: mRNA
A:Residues: 1-997 <HOB>
A:Cross-references: UNIPROT:P53569; UNIPARC:UPI0000029995; GB:U19891; NID:G790574; PIDN:
A:Accession: S65568
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-454, 'KR', 984-985, 'IQA' <HO2>
A:Cross-references: UNIPARC:UPI00000279A2; EMBL:U19892; NID:G790576; PIDN:AAB01504.1; PI
C:Comment: This protein is important in mammalian cells for both growth promoting and gr
C:Genetics:
A:Gene: CBF
A:Map position: 17
C:Keywords: alternative splicing; growth regulation
F:943-967/Region: nuclear location signal

Query Match 42.3%; Score 44; DB 2; Length 997;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLI 15
||| ||||| |||
Db 46 LRLGGTKQDYLM 57

RESULT 15
A36368
transcription factor CBF, CCAAT-binding - human
C:Species: Homo sapiens (man)
C>Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 09-Jul-2004
C:Accession: A36368
R:Lum, L.S.Y.; Sultzman, L.A.; Kaufman, R.J.; Linzer, D.I.H.; Wu, B.J.
Mol. Cell. Biol. 10, 6709-6717, 1990
A:Title: A cloned human CCAAT-box-binding factor stimulates transcription from the human
A:Reference number: A36368; MUID:91061780; PMID:2247079
A:Accession: A36368
A:Molecule type: mRNA
A:Residues: 1-998 <LUM>
A:Cross-references: UNIPROT:Q03701; UNIPARC:UPI0000127180; GB:M37197; NID:G179968; PIDN:
C:Genetics:
A:Gene: GDB:CEBPA; CEBP
A:Cross-references: GDB:128839; OMIM:116897
A:Map position: 19q13.1-19q13.1

Query Match 42.3%; Score 44; DB 2; Length 998;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLI 15
||| ||||| |||
Db 46 LRLGGTKQDYLM 57

RESULT 16
S49346
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Rhodobacter capsulatus
N:Alternate names: cb-type cytochrome-c oxidase 28K chain; ccoO protein; cytochrome b410
C:Species: Rhodobacter capsulatus
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S65859; S4235; D54235; F54235; F54235; S49346
R:Thoeny-Meyer, L.; Beck, C.; Preisig, O.; Hennecke, H.
Mol. Microbiol. 14, 705-716, 1994
A:Title: The ccoNOQP gene cluster codes for a cb-type cytochrome oxidase that functions

A:Reference number: S65858; MUID:95198544; PMID:7891558
A:Accession: S65859
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-242 <THO>
A:Cross-references: UNIPROT:Q52687; UNIPARC:UPI00000B62E9; EMBL:X80134; NID:G556812; PIDN:
R:Gray, K.A.; Grooms, M.; Myllykallio, H.; Moomaw, C.; Slaughtter, C.; Daldal, F.
Biochemistry 33, 3120-3127, 1994
A:Title: Rhodobacter capsulatus contains a novel cb-type cytochrome c oxidase without a c
A:Reference number: A54235; MUID:94176508; PMID:8130227
A:Accession: C54235
A:Molecule type: protein
A:Residues: 'XX', 3-6, 'XX', 9-20 <GRA1>
A:Cross-references: UNIPARC:UPI000017219C
A:Experimental source: PMT0-404/MT-RBC1 cells
A:Note: sequence extracted from NCBI backbone (NCBIP:144512)
A:Accession: D54235
A:Molecule type: protein
A:Residues: 119, 'H', 121-128, 'S', 130-132, 'XSG', 136, 'F' <GRA2>
A:Cross-references: UNIPARC:UPI000017219D
A:Experimental source: PMT0-404/MT-RBC1 cells
A:Note: sequence extracted from NCBI backbone (NCBIP:144514)
A:Accession: E54235
A:Molecule type: protein
A:Residues: 160-164, 'L', 166-170, 'YDAPFQAN' <GRA3>
A:Cross-references: UNIPARC:UPI000017219E
A:Experimental source: PMT0-404/MT-RBC1 cells
A:Note: sequence extracted from NCBI backbone (NCBIP:144517)
A:Accession: F54235
A:Molecule type: protein
A:Residues: 180-192, 'A', 194, 'XAN' <GRA4>
A:Cross-references: UNIPARC:UPI000017219F
A:Experimental source: PMT0-404/MT-RBC1 cells
A:Note: sequence extracted from NCBI backbone (NCBIP:144519)
C:Genetics:
C:Gene: ccoO
C:Function:
A:Description: this cytochrome-c oxidase complex catalyzes the oxidation of four molecules
A:Pathway: respiratory chain
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory c

Query Match 41.3%; Score 43; DB 1; Length 242;
Best Local Similarity 52.9%; Pred. No. 21;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 AYLRFGGTTKDFLI 18
|||: ||| |||
Db 222 AYLRVLTGTVDFSTFQP 238

RESULT 17
B47468
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Bradyrhizobium japonicum
N:Alternate names: cb-type cytochrome-c oxidase 28K chain; cytochrome b410; fixO protein;
C:Species: Bradyrhizobium japonicum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B47468
R:Preisig, O.; Anthamatten, D.; Hennecke, H.
Proc. Natl. Acad. Sci. U.S.A. 90, 3309-3313, 1993
A:Title: Genes for a microaerobically induced oxidase complex in Bradyrhizobium japonicum
A:Reference number: A47468; MUID:93234486; PMID:8386371
A:Accession: B47468
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-244 <PRE>
A:Cross-references: UNIPROT:Q03074; UNIPARC:UPI0000011F85; GB:L07487; NID:G152196; PIDN:
A:Experimental source: 110spc4
A:Note: sequence extracted from NCBI backbone (NCBIP:129654)
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory c

Query Match 41.3%; Score 43; DB 1; Length 244;

C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: A69191
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: A69191
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-240 <MTH>
A:Cross-references: UNIPROT:Q26779; UNIPARC:UPI0000129BAD; GB:AE000848; GB:AE000666; NID
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH683
A:Start codon: GTG
C:Superfamily: tRNA nucleotidyltransferase
C:Keywords: nucleotidyltransferase; tRNA processing

Query Match 40.4%; Score 42; DB 2; Length 240;
Best Local Similarity 38.9%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIFDPK 19
: || || | : : :
Db 43 SYLEFGGKILVAVYGR 60

RESULT 23
H81327
hypothetical protein Cj1214c [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: H81327
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: H81327
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <PAR>
A:Cross-references: UNIPROT:Q9PN81; UNIPARC:UPI00000C1B6C; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1214c

Query Match 40.4%; Score 42; DB 2; Length 241;
Best Local Similarity 47.1%; Pred. No. 30;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFGGKTDFLIFD 17
: || || | : : :
Db 204 PFYSRIGGYKSNFDFN 220

RESULT 24
H72485
probable hydantoinase APE2528 - Aeropyrum pernix (strain KI)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: H72485
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: H72485
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-687 <KAW>
A:Cross-references: UNIPROT:Q9Y8V6; UNIPARC:UPI000005E3A4; DDBJ:AP000064; NID:G5105945;
A:Experimental source: strain KI
C:Genetics:
A:Gene: APE2528
C:Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 40.4%; Score 42; DB 2; Length 687;
Best Local Similarity 63.8%; Pred. No. 90;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GGTGTDLIFD 17
: || || | : : :
Db 14 GGTGTDLIFD 24

RESULT 25
A49672
transcription factor Nrf1 - human
N:Alternate names: basic leucine-zipper transcription factor; NF-E2-related factor 1
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A49672
R:Chan, J.Y.; Han, X.L.; Kan, Y.W.
Proc. Natl. Acad. Sci. U.S.A. 90, 11371-11375, 1993
A:Title: Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in
A:Reference number: A49672; MUID:94068605; PMID:8248256
A:Accession: A49672
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-742 <CHA>
A:Cross-references: UNIPROT:Q14494; UNIPARC:UPI000002AF20
A:Experimental source: erythroleukemia cell line K562
A:Note: sequence extracted from NCBI backbone (NCBIN:140521, NCBIPI:140522)
C:Superfamily: human transcription factor TFC11; fos/jun DNA-binding domain homology
F:618-659/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 40.4%; Score 42; DB 2; Length 742;
Best Local Similarity 53.8%; Pred. No. 97;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 7 GGTGTDLIFDPK 19
: || || | : : :
Db 339 GGCSQDFLLFSPE 351

RESULT 26
A55004
transcription factor TFC11 - human
N:Alternate names: LCR-F1 protein
C:Species: Homo sapiens (man)
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 09-Jul-2004
C:Accession: A55004; S48097
R:Luna, L.; Johnsen, O.; Skartlien, A.H.; Pedetour, F.; Turc-Carel, C.; Prydz, H.; Kolst
Genomics 22, 553-562, 1994
A:Title: Molecular cloning of a putative novel human bZIP transcription factor on chromo
A:Reference number: A55004; MUID:95095252; PMID:8001966
A:Accession: A55004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-772 <LUN>
A:Cross-references: UNIPROT:Q14494; UNIPARC:UPI000012FFCF; GB:X77366; NID:G541677; PIDN:
R:Caterina, J.J.; Donze, D.; Sun, C.W.; Ciavatta, D.J.; Townes, T.M.
Nucleic Acids Res. 22, 2383-2391, 1994
A:Title: Cloning and functional characterization of LCR-F1: a bZIP transcription factor t
A:Reference number: S48097; MUID:94310069; PMID:8036168
A:Accession: S48097
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 326-772 <CAT>
A:Cross-references: UNIPARC:UPI0000016A09C
C:Genetics:
A:Gene: GDB:TCF11

A;Cross-references: GDB:293921; OMIM:600115
A;Map position: 17q22-17q22
C;Superfamily: human transcription factor TFC11; fos/jun DNA-binding domain homology
C;Keywords: DNA binding; leucine zipper; transcription factor
F;648-689/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 40.4%; Score 42; DB 2; Length 772;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 GGTGTDFLIFDPK 19
||| ||| |||
Db 369 GGCQDFLLFSPE 381

RESULT 27
JC4169
phosphoenolpyruvate carboxylase (EC 4.1.1.31) - Thermus sp.
C;Species: Thermus sp.
C;Date: 12-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C;Accession: JC4169; PC4043
R;Nakamura, T.; Yoshioka, I.; Takahashi, M.; Toh, H.; Izui, K.
J. Biochem. 118, 319-324, 1995
A;Title: Cloning and sequence analysis of the gene for phosphoenolpyruvate carboxylase
A;Reference number: JC4169; MUID:96064150; PMID:8543565
A;Contents: No. 71
A;Accession: JC4169
A;Molecule type: DNA
A;Residues: 1-857 <NAK>
A;Cross-references: UNIPARC:UPI0000126F06; DDBJ:D42166; NID:G1061019; PIDN:BAA07723.1; E
A;Accession: PC4043
A;Molecule type: protein
A;Residues: 26-43;120-145;198-212;214-232;270-292;295-319;417-444;801-825 <NA2>
A;Cross-references: UNIPARC:UPI0000175FA8; UNIPARC:UPI0000175FA7; UNIPARC:UPI0000175FA8;
FAD
C;Comment: This enzyme catalyzes the reaction of phosphoenolpyruvate with HCO3-to form o
erotic role by replenishing C4 dicarboxylic acids in the citric acid cycle. In C4 plants
C;Genetics:
A;Gene: ppc
C;Superfamily: phosphoenolpyruvate carboxylase
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 40.4%; Score 42; DB 2; Length 857;
Best Local Similarity 50.08; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLIFDPK 19
||| ||| |||
Db 390 LRLGGVHPDFLALSP 405

RESULT 28
T24474
hypothetical protein T04H1.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T24474
R;Harris, B.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19896
A;Accession: T24474
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-144 <WIL>
A;Cross-references: UNIPROT:Q22176; UNIPARC:UPI000006113D; EMBL:Z78200; PIDN:CAB01580.1;
A;Experimental source: clone T04H1
C;Genetics:
A;Gene: CESP:T04H1.3
A;Map position: 5
A;Introns: 57/1
C;Superfamily: Caenorhabditis hypothetical protein C40H1.5

Query Match 39.4%; Score 41; DB 2; Length 144;
Best Local Similarity 53.2%; Pred. No. 48;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Best Local Similarity 66.7%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 8 GTKTDFLIFDPK 19
||| ||| |||
Db 77 GTYTDILTLDPK 88

RESULT 29
S76494
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S76494
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76494
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-227 <KAN>
A;Cross-references: UNIPROT:P74519; UNIPARC:UPI000012E6FD; EMBL:D90915; GB:AB001339; NID:
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Escherichia coli lipote-protein ligase lipB

Query Match 39.4%; Score 41; DB 2; Length 227;
Best Local Similarity 44.4%; Pred. No. 41;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDP 18
||| ||| ||| |||
Db 47 PVTTLGTGSGTKYLKFPD 64

RESULT 30
S76610
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S76610
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76610
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <KAN>
A;Cross-references: UNIPROT:Q55802; UNIPARC:UPI00000D3563; EMBL:D64004; GB:AB001339; NID:
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: ompR protein; response regulator homology
C;Keywords: phosphoprotein
F;33-149/Domain: response regulator homology <RRH>
F;88/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 39.4%; Score 41; DB 1; Length 262;
Best Local Similarity 56.2%; Pred. No. 48;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLIFDPK 19
||| ||| |||
Db 103 LRFGGDNTFILLIVSAK 118

RESULT 31
A72209

hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1998 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

C:Accession: A72209

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Thermotoga maritima*

A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: A72209

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-444 <ARN>

A:Cross-references: UNIPROT:Q9X2D2; UNIPARC:UPI00000C11ED; GB:AE000512; NID:10360571

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1812

Query Match 39.4%; Score 41; DB 2; Length 444;

Best Local Similarity 66.7%; Pred. No. 83;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 YLRFGGTKTDFL 14

Db 226 YLMSGGKSDFL 237

||| ||| ||| |||

||| ||| ||| |||

RESULT 32

A86303

hypothetical protein F17F16.4 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: A86303

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chao, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.;

anssen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; PMID:21016719; PMID:11130712

A:Accession: A86303

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-471 <STO>

A:Cross-references: UNIPROT:Q9FWQ9; UNIPARC:UPI00000A3269; GB:AE005172; NID:g9954731; PI

C:Genetics:

A:Map position: 1

Query Match 39.4%; Score 41; DB 2; Length 471;

Best Local Similarity 53.3%; Pred. No. 88;

Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTKTDFLIF 16

Db 278 AYLAIVGPKTDLKULF 292

||| ||| ||| |||

||| ||| ||| |||

RESULT 33

A70362

N-methylhydantoinase A - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004

C:Accession: A70362

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70300; PMID:98196666; PMID:9537320

A:Accession: A70362

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-660 <AQF>

A:Cross-references: UNIPROT:O66925; UNIPARC:UPI0000056417; GB:AE000703; NID:g2983287; PI

A:Experimental source: strain VF5

C:Genetics:

A:Gene: hyuA

C:Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 39.4%; Score 41; DB 2; Length 660;

Best Local Similarity 61.5%; Pred. No. 1.3e+02;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 GGTCTDFLIFDPK 19

Db 10 GGTFTDFVWDGK 22

||| ||| ||| |||

||| ||| ||| |||

RESULT 34

B72608

probable hyuA APE1328 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C:Accession: B72608

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahashi,

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kikuchi,

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum*

A:Reference number: A72450; PMID:99310339; PMID:10382966

A:Accession: B72608

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-785 <KAW>

A:Cross-references: UNIPROT:Q9YCC8; UNIPARC:UPI000005DEDD; DDBJ:AP000061; NID:G5104821; I

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1328

C:Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 39.4%; Score 41; DB 2; Length 785;

Best Local Similarity 63.6%; Pred. No. 1.5e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GGTCTDFLIFD 17

Db 116 GGTFTDIVVFD 126

||| ||| ||| |||

||| ||| ||| |||

RESULT 35

T17464

rifamycin polyketide synthase modules 4-6 - *Amycolatopsis mediterranei*

C:Species: *Amycolatopsis mediterranei*

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004

C:Accession: T17464

R:Schupp, T.

submitted to the EMBL Data Library, December 1997

A:Reference number: Z18802

A:Accession: T17464

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-5069 <SCH>

A:Cross-references: UNIPROT:O52789; UNIPARC:UPI0000055B02; EMBL:AJ223012; NID:e1227119; I

A:Experimental source: strain LBG A3136

C:Keywords: carrier protein

F1631-1702/Domain: acyl carrier protein homology <ACP1>

F3238-3309/Domain: acyl carrier protein homology <ACP2>

F4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 39.4%; Score 41; DB 2; Length 5069;

Best Local Similarity 47.1%; Pred. No. 1e+03;

Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTKDPLIFDP 18
||:||||:||||
Db 1783 AYLRQGGFLHEALFDP 1799

RESULT 36

B44827
FMRamide-like peptide - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: B44827
R:Rosoff, M.L.; Burglin, T.R.; Li, C.
J. Neurosci. 12, 2356-2361, 1992
A:Title: Alternatively spliced transcripts of the flp-1 gene encode distinct FMRamide-1
A:Reference number: A44827; MUID:92300457; PMID:1607945
A:Accession: B44827
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-164 <ROS>
A:Cross-references: UNIPROT:P41855; UNIPARC:UPI000017B6C7
A:Note: sequence extracted from NCBI backbone (NCBIP:106594)
C:Keywords: neuropeptide

Query Match 38.9%; Score 40.5; DB 2; Length 164;
Best Local Similarity 43.5%; Pred. No. 36;
Matches 10; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

QY 1 PAVLRFGGKTKDPLIF-----DP 18
||:||||:||||
Db 104 PNFILRFGSRDPNPLRFGKAAADP 126

RESULT 37

G98278
hypothetical protein AGR_L_2370 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G98278
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G98278
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-99 <KUR>
A:Cross-references: UNIPROT:Q8U4W7; UNIPARC:UPI00000D2931; GB:AE007870; PIDN:AAK89753.1;
C:Genetics:
A:Gene: AGR_L_2370
A:Map position: linear chromosome

Query Match 38.5%; Score 40; DB 2; Length 99;
Best Local Similarity 44.4%; Pred. No. 25;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 PAVLRFGGKTKDPLIFDP 18
||||:||||:
Db 8 PAYLIQGGKLDITLAIP 25

RESULT 38

T51959
hypothetical protein [imported] - Picea mariana
C:Species: Picea mariana
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004
C:Accession: T51959
R:Perry, D.J.; Bousquet, J.
Genetics 149, 1089-1098, 1998
A:Title: Sequence-tagged-site (STS) markers of arbitrary genes. Development, characteriz
A:Reference number: Z25268; MUID:98278823; PMID:9611216
A:Accession: T51959
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-141 <PER>

A:Cross-references: UNIPROT:O65048; UNIPARC:UPI00000A2B06; EMBL:AF051205; PIDN:AAC32110.1

A:Experimental source: embryo

C:Genetics:

A:Note: Sb08

Query Match 38.5%; Score 40; DB 2; Length 141;
Best Local Similarity 58.8%; Pred. No. 37;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 4 LRFG--GTTKDFLIFDP 18
||||:||||:
Db 19 LRFGLAGVKSILISHP 35

RESULT 39

AD1156
hypothetical protein lmo0652 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AD1156

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001

A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,
A:Title: Comparative genomics of Listeria species.

A:Reference number: ABI077; MUID:21537279; PMID:11679669

A:Accession: AD1156

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <GLA>

A:Cross-references: UNIPROT:O8Y981; UNIPARC:UPI0000055170; GB:NC_003210; PIDN:CAC98730.1;

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0652

Query Match 38.5%; Score 40; DB 2; Length 178;
Best Local Similarity 42.9%; Pred. No. 47;
Matches 9; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 3 YLRFG---GTTKDFLIFDPK 19
||||:||||:
Db 154 YLRGQFLGTGNGFHLVQKK 174

RESULT 40

AB2204

lipote-protein ligase B [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C:Accession: AB2204

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2204

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-221 <KUR>

A:Cross-references: UNIPROT:Q8YSA4; UNIPARC:UPI000012E6B5; GB:BA0000019; PIDN:BA074884.1;

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3185

C:Superfamily: Escherichia coli lipote-protein ligase lipB

Query Match 38.5%; Score 40; DB 2; Length 221;
Best Local Similarity 44.4%; Pred. No. 59;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

```
Qy 1 PAYLRFGGTKTDFLIFDP 18
      | | | | |
      : : : : :
Db 53 PVTTLGQSSLEFLAKFDP 70

RESULT 41
AG3447
cytochrome-c oxidase (EC 1.9.3.1) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AG3447
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: ADJ252; PMID:11756688
A:Accession: AG3447
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <KUR>
A:Cross-references: UNIPROT:Q8YFF7; UNIPROT:Q8G2G0; UNIPARC:UPI0000058103; GB:AE008917;
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11565
A:Map position: 1
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C:Keywords: oxidoreductase

Query Match 38.5%; Score 40; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 10; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

Qy 2 AYLRFGGKTDFLIFDP--PK 19
      | | | | |
      : : : : :
Db 222 AYLQMLGLTVDFSTYDQSPK 241

RESULT 42
H97341
dihydrodipicolinate synthase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97341
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: H97341
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <KUR>
A:Cross-references: UNIPROT:Q97D80; UNIPARC:UPI0000128535; GB:AE001437; PIDN:AAK81523.1;
A:Experimental source: Clostridium acetobutylicum ATCC924
C:Genetics:
A:Gene: CAC3600
C:Superfamily: dihydrodipicolinate synthase

Query Match 38.5%; Score 40; DB 2; Length 286;
Best Local Similarity 42.1%; Pred. No. 77;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19
      | | | | |
      : : : : :
Db 72 PIYVFGGNNTKMTKD1K 90

RESULT 43
T35417
probable beta-lactamase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T35417
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21577
A:Accession: T35417
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-310 <OLI>
A:Cross-references: UNIPROT:Q9X7X2; UNIPARC:UPI000000DAF52; EMBL:AL049485; PIDN:CAB39710.1
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC6A5.26c

Query Match 38.5%; Score 40; DB 2; Length 310;
Best Local Similarity 57.1%; Pred. No. 83;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFL 14
      | | | | |
      : : : : :
Db 86 PQIPQAGLKTDFL 99

RESULT 44
D89102
protein F25E5.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89102
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_eleg
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D89102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <STO>
A:Cross-references: UNIPROT:O76655; UNIPARC:UPI000007D193; GB:chr_V; PIDN:AAC27333.1; PI
C:Genetics:
A:Gene: F25E5.2
A:Map position: 5

Query Match 38.5%; Score 40; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YLRFGGTKTDFLIF 16
      | | | | |
      : : : : :
Db 257 FLKFGSKEDLAKF 270

RESULT 45
AC0382
probable Pyridoxal-phosphate dependent protein YP03147 [imported] - Yersinia pestis (str
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC0382
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AC0382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <KUR>
A:Cross-references: UNIPROT:Q8ZC75; UNIPARC:UPI00000DC7C3; GB:AL590842; PIDN:CAC92382.1;
C:Genetics:
A:Gene: YP03147

Query Match 38.5%; Score 40; DB 2; Length 346;
```


C;Accession: G71657
R;Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: G71657
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-433 <AND>
A;Cross-references: UNIPROT:Q9ZD14; UNIPARC:UPI00000D37C8; GB:AJ235272; GB:AJ235269; NID
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: folC; RP536
C;Superfamily: folypolyglutamate synthase

Query Match	38.5%	Score 40;	DB 2;	Length 433;
Best Local Similarity	50.0%;	Pred. No. 1.2e+02;		
Matches	8;	Conservative	2;	Mismatches 6; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTKDFLIPD 17
|:|:|:|:|:|:|:
Db 121 AFLAPAQTKADILILE 136

Search completed: June 5, 2006, 12:53:28
Job time : 39.0822 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM protein - protein search, using sw model
Run on: June 5, 2006, 12:32:17 ; Search time 166.575 Seconds
(without alignments)
105.510 Million cell updates/sec

Title: US-10-645-659A-7
Perfect score: 104
Sequence: 1 PAYLRFQGTGTDLFIDPK 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 1000 summaries

Database : UniProt 7.2.*
1: uniprot_sprot.*
2: uniprot_crembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	543	1 HPSE_HUMAN	Q9Y251 homo sapien
2	104	100.0	558	2 Q33X5 SPAJD	Q33X5 spalax juda
3	104	100.0	574	2 Q33X6 SPAJD	Q33X6 spalax juda
4	104	100.0	574	2 Q33X7 GRODE	Q33X7 spalax carm
5	104	100.0	574	2 Q33X8 GRODE	Q33X8 spalax gola
6	104	100.0	574	2 Q33X9 GRODE	Q33X9 spalax gali
7	99	95.2	535	1 HPSE_MOUSE	Q6Y9Z1 mus musculus
8	99	95.2	536	1 HPSE_RAT	Q71RP1 rattus norv
9	92	88.5	545	1 HPSE_BOVIN	Q9MYV0 bos taurus
10	81	77.9	523	1 HPSE_CHICK	Q9DYK5 gallus gall
11	75	72.1	533	2 Q4SYF6 TETNG	Q4SYF6 tetraodon n
12	66	63.5	592	1 HPSE2_HUMAN	Q8WWQ2 homo sapien
13	66	63.5	592	2 Q2M1H9 HUMAN	Q2M1H9 homo sapien
14	66	63.5	597	2 Q4TB80 TETNG	Q4TB80 tetraodon n
15	53	51.0	326	2 Q7MH05 VIBVY	Q7MH05 vibrio vuln
16	53	51.0	326	2 Q8DCU6 VIBVU	Q8DCU6 vibrio vuln
17	53	51.0	1538	2 Q2QP6 LEPIC	Q2QP6 leptospira
18	53	51.0	1538	2 Q8F55 LEPIN	Q8F55 leptospira
19	50	48.1	512	2 Q8LHX2 ORYSA	Q8LHX2 oryza sativ
20	49.5	47.6	224	2 Q4HLV4 CAMLA	Q4HLV4 campylobact
21	49	47.1	513	2 Q6Z4S8 ORYSA	Q6Z4S8 oryza sativ
22	49	47.1	588	2 Q67183 AQUAE	Q67183 aquifex aeo
23	48	46.2	161	2 Q8PS30 METMA	Q8PS30 methanosarc
24	48	46.2	176	2 Q8TKK5 METAC	Q8TKK5 methanosarc
25	48	46.2	194	2 Q2J1Z0 9CYAN	Q2J1Z0 cyanobacter
26	48	46.2	208	2 Q86YU0 HUMAN	Q86YU0 homo sapien
27	48	46.2	247	2 Q2RP20 RHORU	Q2RP20 rhodospiril
28	48	46.2	264	1 CAH7_HUMAN	P43166 homo sapien
29	48	46.2	265	2 Q9Z8Y1 CHLPN	Q9Z8Y1 chlamydia p
30	48	46.2	311	2 Q41BD7 9BACI	Q41BD7 exigubacte
31	48	46.2	411	2 Q8G431_BIFLO	Q8G431 bifidobacte

32	48	46.2	425	2	Q3FBC5_9BURK	Q3fbc5 burkholderi
33	48	46.2	515	2	Q8T108_BOMMO	Q8t108 bombyx mori
34	48	46.2	638	2	Q47YD5_COLP3	Q47yds colwallia p
35	48	46.2	742	2	Q8KZS8_ACEPA	Q8kzs8 acetobacter
36	48	46.2	838	2	Q2IRU0_RHOPA	Q2iru0 rhodopseudo
37	48	46.2	1895	2	Q3QRL3_9RHOB	Q3qrl3 silicibacte
38	48	46.2	2405	2	Q2LZG0_DROPS	Q2lzo0 drosophila
39	47	45.2	176	2	Q9WZB4_THEMA	Q9wzb4 thermotoga
40	47	45.2	1076	2	Q5ZIQ7_CHICK	Q5ziq7 gallus gall
41	47	45.2	1076	2	Q5ZMV4_CHICK	Q5zmv4 gallus gall
42	46	44.2	180	2	Q33FF2_METHU	Q33ff2 methanospir
43	46	44.2	219	2	Q4CAE2_CROWT	Q4cae2 crocosphaer
44	46	44.2	308	2	Q5TMQ4_ANOGA	Q5tmq4 anopheles g
45	46	44.2	319	2	Q2W843_MAGSA	Q2w843 magnetospir
46	46	44.2	319	2	Q2WAD9_MAGSA	Q2wad9 magnetospir
47	46	44.2	404	2	Q57MP2_SALCH	Q57mp2 salmonella
48	46	44.2	404	2	Q5PDY6_SALPA	Q5pdy6 salmonella
49	46	44.2	404	2	Q9F7A7_SALTY	Q9f7a7 salmonella
50	46	44.2	405	1	WCAD_ECOLI	P71238 escherichia
51	46	44.2	405	2	Q323F6_SHIBS	Q323f6 shigella bo
52	46	44.2	405	2	Q32OE7_SHISS	Q32oe7 shigella so
53	46	44.2	405	2	Q2MAY3_ECOLI	Q2may3 escherichia
54	46	44.2	405	2	Q8X7M9_ECO57	Q8x7m9 escherichia
55	46	44.2	414	2	Q7XIE4_ORYSA	Q7xie4 oryza sativ
56	46	44.2	429	2	Q8FG19_ECOL6	Q8fg19 escherichia
57	46	44.2	445	2	Q9ANW0_ECO57	Q9anw0 escherichia
58	46	44.2	507	2	Q8LHX1_ORYSA	Q8lhx1 oryza sativ
59	46	44.2	527	2	Q9LRC8_SCUBA	Q9lrc8 acutellaria
60	46	44.2	536	2	Q2UDS9_ASPOR	Q2uds9 aspergillus
61	46	44.2	1245	2	Q9YGH8_SCOMX	Q9ygh8 scophthalmu
62	45.5	43.8	251	1	SFSA_SYMTH	Q67GV9 symbiobacte
63	45.5	43.8	385	2	Q5S2S8_DICDI	Q5s2s8 dictyosteli
64	45.5	43.8	385	2	Q861C3_DICDI	Q861c3 dictyosteli
65	45.5	43.8	486	2	Q3DX84_CHLAU	Q3dx84 chloroflexu
66	45	43.3	183	2	Q64Z07_BACFR	Q64z07 bacteroides
67	45	43.3	212	2	Q8BIS4_MOUSE	Q8bis4 mus musculu
68	45	43.3	241	2	Q3PHF0_PARDE	Q3phf0 paracoccus
69	45	43.3	241	2	Q51680_PARDE	Q51680 paracoccus
70	45	43.3	246	2	Q43943_AZOCA	Q43943 azorhizobiu
71	45	43.3	250	2	Q942D6_ORYSA	Q942d6 oryza sativ
72	45	43.3	451	2	Q8ESC3_OCBIH	Q8esc3 oceanobacil
73	45	43.3	541	2	Q691I6_ORYSA	Q691i6 oryza sativ
74	45	43.3	557	2	Q4J5V6_AZOVI	Q4j5v6 azotobacter
75	45	43.3	591	2	Q2W0L1_MAGSA	Q2w0l1 magnetospir
76	45	43.3	1006	2	Q5U4R2_XENLA	Q5u4r2 xenopus lae
77	45	43.3	1096	2	Q616F9_CABER	Q616f9 caenorhabdi
78	45	43.3	1623	2	Q8K449_MOUSE	Q8k449 mus musculu
79	45	43.3	3842	2	Q7R0N8_GIALA	Q7r0n8 giardia lam
80	44.5	42.8	606	2	Q2N8C9_9SPHN	Q2n8c9 erythroba
81	44.5	42.8	750	2	Q3RSB0_RALME	Q3rsb0 ralstonia m
82	44	42.3	136	2	Q3UUZ3_MOUSE	Q3uu3 mus musculu
83	44	42.3	190	2	Q979H0_THEVO	Q979h0 thermoplasm
84	44	42.3	284	2	Q9HNE6_HALSA	Q9hne6 halobacteri
85	44	42.3	285	2	Q46JDI_PROMT	Q46jdi prochloroto
86	44	42.3	303	2	Q81QF9_BACAN	Q81qf9 bacillus an
87	44	42.3	305	2	Q6F2A4_MESFL	Q6f2a4 mesoplaema
88	44	42.3	435	2	Q8K2R7_MOUSE	Q8k2r7 mus musculu
89	44	42.3	536	2	Q2UT15_ASPOR	Q2ut15 aspergillus
90	44	42.3	568	2	Q5ZKA0_CHICK	Q5zka0 gallus gall
91	44	42.3	606	2	Q8LHW4_ORYSA	Q8lhw4 oryza sativ
92	44	42.3	652	2	Q4E3C0_TRYCR	Q4e3c0 trypanosoma
93	44	42.3	762	2	Q9A8V8_CAUCR	Q9a8v8 caulobacter
94	44	42.3	772	2	Q5RA25_PONPY	Q5ra25 pongo pygma
95	44	42.3	961	2	Q3UWQ7_MOUSE	Q3uwq7 mus musculu
96	44	42.3	997	1	CEBPZ_MOUSE	P53559 mus musculu
97	44	42.3	998	1	CEBPZ_HUMAN	Q33701 homo sapien
98	44	42.3	1002	2	Q2JQ83_9CYAN	Q2jq83 cyanobacter
99	44	42.3	1040	2	Q3TXW5_MOUSE	Q3txw5 mus musculu
100	44	42.3	1052	2	Q3TYA8_MOUSE	Q3tya8 mus musculu
101	44	42.3	1054	2	Q8NE75_HUMAN	Q8ne75 homo sapien
102	44	42.3	1060	2	Q7SFT5_NEUCR	Q7sf55 neurospora
103	44	42.3	1369	2	Q8UW86_PAROL	Q8uw86 paralichthy
104	44	42.3	2049	2	Q5W6T2_ORYSA	Q5w6t2 oryza sativ

981 38 36.5 330 2 Q41UT3_FERAC Q41ut3 ferroplasma
 982 38 36.5 332 2 Q6LM33_PROPR Q6lm33 photobacter
 983 38 36.5 333 2 Q2P49_RHIME Q2p49 rhizobium m
 984 38 36.5 334 2 Q2J5W4_9ACTO Q2j5w4 frankia sp.
 985 38 36.5 335 2 Q3NYA7_9CAGM Q3nya7 shewanella
 986 38 36.5 338 2 Q4R6K4_MACFA Q4r6k4 macaca fasc
 987 38 36.5 338 2 Q407C2_JANNASCHIA Q407c2 jannaschia
 988 38 36.5 338 2 Q9WZQ3_THERMOTOGA Q9wzq3 thermotoga
 989 38 36.5 341 1 P04970_CAENORHABDI P04970 caenorhabdi
 990 38 36.5 341 1 G3P4_CAEEL P17331 caenorhabdi
 991 38 36.5 341 2 Q4FSE8_PSYAR Q4fse8 psychrobact
 992 38 36.5 342 2 Q43Q84_SOLUS Q43q84 solibacter
 993 38 36.5 343 2 Q46823_TRIASC Q46823 triakis ecy
 994 38 36.5 344 2 Q5FTK8_GLUOX Q5ftk8 gluconobact
 995 38 36.5 346 2 Q9SHG1_ARATH Q9shg1 arabidopsis
 996 38 36.5 349 2 Q3E8L3_ARATH Q3e8l3 arabidopsis
 997 38 36.5 350 2 Q6J678_9BURK Q6j678 collimonas
 998 38 36.5 353 2 Q08563_RAT Q08563 rattus norv
 999 38 36.5 354 2 Q4JV13_CORJK Q4jv13 corynebacte
 1000 38 36.5 355 2 Q37X36_NOVOSHINGO Q37x36 novoshing

ALIGNMENTS

RESULT 1

HPSE HUMAN STANDARD; PRT; 543 AA.
 ID Q9Y251; Q53GB5; Q9UL39;
 AC 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-NOV-1999, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Heparanase precursor (PC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-
 DE glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
 DE subunit).
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRI, HPSE1, HSE1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
 RA Seddon A.P., Giorgio N.A., Bohlen P.;
 RT "Cloning and functional expression of a human heparanase gene.";
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.
 RC TISSUE=Embryonic fibroblast;
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
 RA Toyoshima M., Nakajima M.;
 RT "Human heparanase. Purification, characterization, cloning, and
 RT expression.";
 RL J. Biol. Chem. 274:24153-24160(1999).
 RN [3]
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
 RX PubMed=10395325; DOI=10.1038/10518;
 RA Vlodavsky I., Friedmann Y., Elkin M., Angorn H., Atzmon R.,
 RA Izhai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
 RA Spector L., Pecker I.;
 RT "Mammalian heparanase: gene cloning, expression and function in tumor
 RT progression and metastasis.";
 RL Nat. Med. 5:793-802(1999).
 RN [4]
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 RN [5]
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage.";
 RL Glycobiology 10:467-475(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RX PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis.";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
 RN [7]
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Farek R., Stubbsfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpal) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435(2003).
 RN [8]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RA Pinnal M.A., Semedo P.;
 RT "Cloned heparanase from MCF-7 cells.";
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Small intestine;
 RX Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [11]
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RX PubMed=11123890; DOI=10.1021/bi002080p;
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Gready J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic
RL endoglycosidase heparanase.";
RN Biochemistry 39:15659-15667(2000).
[12]
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
RX ASN-217; ASN-238 AND ASN-459.
RP PubMed=14573609; DOI=10.1074/jbc.M300541200;
RA Shimizu S., Iehida K., Wierzbicka M.K., Osada H.;
RT "Secretion of heparanase protein is regulated by glycosylation in
RL human tumor cell lines.";
RN J. Biol. Chem. 279:2697-2703(2004).
[13]
RP SUBCELLULAR LOCATION.
RX PubMed=15292202; DOI=10.1074/jbc.M402131200;
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
RT "Heparanase uptake is mediated by cell membrane heparan sulfate
RL proteoglycans.";
RN J. Biol. Chem. 279:44084-44092(2004).
[14]
RP SUBCELLULAR LOCATION, PROCESSING, AND SUBCELLULAR LOCATION.
RX PubMed=15948168; DOI=10.1016/j.febslet.2005.03.030;
RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
RT "Heparanase processing by lysosomal/endosomal preparation.";
RL FEBS Lett. 579:2334-2338(2005).
[15]
RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
RA Aboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
RA Elgavish S., Peretz T., Vlodavsky I.;
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
RL human proheparanase.";
RN J. Biol. Chem. 280:13568-13575(2005).
[16]
RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;
RA Levy-Adam F., Aboud-Jarrous G., Guerrini M., Beccati D.,
RA Vlodavsky I., Ilan N.;
RT "Identification and characterization of heparin/heparan sulfate
RL binding domains of the endoglycosidase heparanase.";
RN J. Biol. Chem. 280:20457-20466(2005).
[17]
RP VARIANT SER-260.
RX PubMed=15334672;
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
RA Shi L.H., Shen H.W., Xu B.;
RT "Heparanase mRNA expression and point mutation in hepatocellular
RL carcinoma.";
RN World J. Gastroenterol. 10:2795-2799(2004).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -!- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
CC lower extent, by heparin and sulfamin and activated by calcium and
CC magnesium (By similarity).
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
CC pH dependence:
CC Optimum pH is 4-6;
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products.
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium.
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
CC weakly expressed in lymph node, thymus, peripheral blood
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme.
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.
Query Match 100.0%; Score 104; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 9.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKDTFLIFDPK 19
DB 89 PAYLRFGGTKDTFLIFDPK 107
RESULT 2
Q333X5 SPAJD
ID Q333X5_SPAJD PRELIMINARY; PRT; 558 AA.
AC Q333X5;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RL cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RL cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AM085494; CAJ30021.1; -; mRNA.
SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F55849EE7 CRC64;
Query Match 100.0%; Score 104; DB 2; Length 558;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKDTFLIFDPK 19
DB 120 PAYLRFGGTKDTFLIFDPK 138
RESULT 3
Q333X6 SPAJD
ID Q333X6_SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=134510;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085493; CAJ30020.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db 120 PAYLRFGGTKTDFLIFDPK 138  
|||||  
-----  
RESULT 4  
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X7_9RODE PRELIMINARY; PRT; 574 AA.  
AC Q333X7;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax carmeli.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=164324;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085492; CAJ30019.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCBBD99DE CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db 120 PAYLRFGGTKTDFLIFDPK 138  
|||||  
-----  
RESULT 5  
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
```

```
AC Q333X8;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax golani.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=191382;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085491; CAJ30018.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64555 MW; 48BEBFECD7D0BCB34 CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db 120 PAYLRFGGTKTDFLIFDPK 138  
|||||  
-----  
RESULT 6  
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X9_9RODE PRELIMINARY; PRT; 574 AA.  
AC Q333X9;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax galili.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=164323;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----
```


DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match 100.0%; Score 104; DB 2; Length 574;
Best Local Similarity 100.0%; Pred. NO. 1e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PAYLRFGGTKDTFLIFDPK 19
|||||
DB 120 PAYLRFGGTKDTFLIFDPK 138

RESULT 7
HPSE MOUSE
ID HPSE MOUSE STANDARD; PRT; 535 AA.
AC Q6VGV21; Q8K3K3;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (EC 3.2.2.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=Hpse; Synonyms=Hpa;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=SJL/J; TISSUE=Spleen;
RX MEDLINE=93121249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND
RP SUBUNITS.
RC STRAIN=FVB; TISSUE=Embryo;
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
RT "Cloning, expression, and purification of mouse heparanase.";
RL Protein Expr. Purif. 26:425-431(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,
RA Li J.-P.;
RT "Processing of macromolecular heparin by heparanase.";
RL J. Biol. Chem. 278:35152-35158(2003).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humniecek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,

Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
Petkovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
Schombach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
Shibata Y., Shimada H., Shinada K., Silva D., Sinclair B.,
Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
Wahlstedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and
CC magnesium (By similarity). Inhibited by laminarin sulfate and, to
CC a lower extent, by heparin and sulfamin.
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
CC Optimum pH is 5;
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products.
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL; AF359507; AAQ15188.1; -; mRNA.
CC EMBL; AY077467; AAL76083.1; -; mRNA.
CC EMBL; AI151051; AAN41636.1; -; mRNA.
CC EMBL; AK040471; BAC30600.1; -; mRNA.
CC EMBL; AK154628; BAE32725.1; -; mRNA.
CC Ensemble; ENSMUSG0000035273; Mus musculus.
CC MGI; MGI:1343124; Hpa.
CC GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
CC InterPro; IPR005199; Glyco_hydro_79_N.
CC Pfam; PF03662; Glyco_hydro_79n; 1.
CC Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
CC Magnesium; Membrane; Signal.
CC SIGNAL 1 27 By similarity.
CC CHAIN 28 101 Heparanase 8 kDa subunit.
CC PROPEP 102 149 /FTRIDPRO.000004263.
CC Linker peptide (By similarity).

```

FT CHAIN 150 535 /FTid=PRO_0000042264.
FT Heparanase 50 kDa subunit.
FT FTid=PRO_0000042265.
FT Heparin/HS-binding (By similarity).
FT REGION 262 272 Heparin/HS-binding (By similarity).
FT REGION 217 217 Proton donor (Potential).
FT ACT_SITE 335 335 Nucleophile (Potential).
FT ACT_SITE 154 154 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 192 192 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 209 209 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 230 230 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 451 451 N-linked (GlcNAc. . .) (Potential).
FT CONFLICT 206 206 K -> R (in Ref. 3).
FT CONFLICT 212 212 W -> S (in Ref. 3).
FT CONFLICT 230 232 NGS -> DGL (in Ref. 1, 2 and 4).
FT CONFLICT 335 335 E -> K (in Ref. 3).
FT CONFLICT 342 342 G -> A (in Ref. 3).
FT CONFLICT 455 455 Y -> H (in Ref. 1, 2 and 4).
FT CONFLICT 531 531 V -> I (in Ref. 1, 2 and 4).
FT CONFLICT 535 535 AA; 60050 MW; AF19828B7CD03F7B CRC64;
SQ SEQUENCE 535 AA; 60050 MW; AF19828B7CD03F7B CRC64;

Query Match 95.2%; Score 99; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 6.9e-08; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 PAYLRFGGTKTDFLI FDP 18
Db 81 PAYLRFGGTKTDFLI FDP 98

RESULT 8
HPSE RAT
ID HPSE RAT STANDARD; PRT; 536 AA.
AC Q1RPL; Q9QZF8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
DE Name=Hpse; Synonyms=Hep;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RP TISSUE=Placenta;
RX MEDLINE=22194309; PubMed=10395326; DOI=10.1038/10525;
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line.";
RL J. Biol. Chem. 277:32459-32465(2002).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes

```

```

CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF359508; AAQ15189.1; -; mRNA.
DR EMBL; AF184967; AAF04563.1; -; mRNA.
DR RGD; 61969; Hpse.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 Linker peptide (By similarity).
FT CHAIN 151 536 /FTid=PRO_0000042267.
FT REGION 151 155 Heparanase 50 kDa subunit.
FT REGION 263 273 /FTid=PRO_0000042268.
FT ACT_SITE 218 218 Heparin/HS-binding (By similarity).
FT ACT_SITE 336 336 Proton donor (Potential).
FT CARBOHYD 155 155 Nucleophile (Potential).
FT CARBOHYD 193 193 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc. . .) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 95.2%; Score 99; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 6.9e-08; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 PAYLRFGGTKTDFLI FDP 18
Db 82 PAYLRFGGTKTDFLI FDP 99

RESULT 9
HPSE BOVIN
ID HPSE BOVIN STANDARD; PRT; 545 AA.
AC Q9MYI0;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
DE Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RP TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;

```

RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
CC EMBL; AF281160; AAF87301.2; -; mRNA.
CC InterPro; IPR005199; Glyco_hydro_79_N.
CC Pfam; PF03662; Glyco_hydro_79n; 1.
CC Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
CC Signal.
CC SIGNAL 1 37 By similarity.
CC CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
CC PROPEP 112 159 /FTID-PRO 0000042256.
CC CHAIN 160 545 /FTID-PRO 0000042257.
CC similarity).
CC /FTID-PRO 0000042258.
CC Heparin/HS-binding (Potential).
CC Heparin/HS-binding (Potential).
CC Proton donor (Potential).
CC Nucleophile (Potential).
CC N-linked (GlcNAc...) (Potential).
CC N-linked (GlcNAc...) (Potential).
CC N-linked (GlcNAc...) (Potential).
CC CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).
CC SEQUENCE 545 AA; 61077 MW; FAC4BDFD85B933 CRC64;
Query Match 88.5%; Score 92; DB 1; Length 545;
Best Local Similarity 89.5%; Pred. No. 1.1e-06;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDPK 19
Db 91 PAYLRFGGKNGDFLIFDPK 109
RESULT 10
HPSE_CHICK STANDARD; PRT; 523 AA.
AC Q90YK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.

DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Names=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=2136959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldehmldt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
RT "Expression pattern and secretion of human and chicken heparanase are
RT determined by their signal peptide sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC (By similarity).
CC -!- PTM: N-glycosylated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
CC EMBL; AY037007; AAK82648.1; -; mRNA.
CC Ensembl; ENSGALG00000011203; Gallus gallus.
CC InterPro; IPR005199; Glyco_hydro_79_N.
CC Pfam; PF03662; Glyco_hydro_79n; 1.
CC Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
CC SIGNAL 1 18 Potential.
CC CHAIN 19 523 Heparanase.
CC /FTID-PRO 0000042259.
CC Heparin/HS-binding (By similarity).
CC Heparin/HS-binding (By similarity).
CC Proton donor (Potential).
CC Nucleophile (Potential).
CC N-linked (GlcNAc...) (Potential).
CC N-linked (GlcNAc...) (Potential).
CC N-linked (GlcNAc...) (Potential).
CC CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
CC CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
CC SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;
Query Match 77.9%; Score 81; DB 1; Length 523;
Best Local Similarity 77.8%; Pred. No. 8.3e-05;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 PAYLRFGGTKTDFLIFDP 18
Db 69 PGFLRFGGTSTDFLIFNP 86
RESULT 11
Q4SYR6 TETNG
ID Q4SYR6 TETNG PRELIMINARY; PRT; 533 AA.
AC Q4SYR6;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE (Fragment).
DE ORFNames=GSTENG00010356001;
GN Tetraodon nigroviridis (Green puffer).
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.

DR EMBL; AL590036; CAH70448.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH16472.1; -; Genomic_DNA.
 DR EMBL; AL139243; CAH16472.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH16472.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH16472.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH16472.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH17160.1; -; Genomic_DNA.
 DR EMBL; AL139243; CAH17160.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH17160.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH17160.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH17160.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH73139.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH73139.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH70450.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH70450.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH70450.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH70450.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH70450.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH16474.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH16474.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH14148.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH14148.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH14148.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH14148.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH14148.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH73138.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH73138.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH73138.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH73138.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH73138.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH17161.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH17161.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH17161.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH17161.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH17161.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH70449.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH70449.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH70449.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH70449.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH70449.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH16473.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH16473.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH16473.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH16473.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH16473.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH14147.1; -; Genomic_DNA.
 DR EMBL; AL356220; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL139243; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL356220; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL356268; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL445251; CAH14147.1; JOINED; Genomic_DNA.
 DR EMBL; AL590036; CAH14147.1; JOINED; Genomic_DNA.
 DR PIR; JC7506; JC7506.
 DR Ensembl; ENSG00000172987; Homo sapiens.
 DR HGNC; HGNC:18374; HPSE2.
 DR GO; GO:0005622; C:intracellular; TAS.
 DR GO; GO:0030305; F:heparanase activity; TAS.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 DR KW Alternative splicing; Hydrolase; Membrane; Polymorphism.
 DR CHAIN 1 592 Heparanase-2.
 DR FT ACT SITE 262 262
 DR FT VARSPLIC 150 261
 DR FT Missing (in isoform 4).
 DR FT

FT VARSPLIC 204 261
 FT Missing (in isoform 3).
 FT FTid=VSP 015851.
 FT SVOLNQPVLV -> TORCQVCGII (in isoform 2).
 FT FTid=VSP 015852.
 FT Missing (in isoform 2).
 FT FTid=VSP 015853.
 FT F -> Y (in dbSNP:10883100).
 FT FTid=VAR 023601.
 FT P -> L (in Ref. 2; CAC82492).
 FT CONFLICT 12 12
 FT CONFLICT 213 213 F -> S (in Ref. 2).
 SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;
 Query Match 63.5%; Score 66; DB 1; Length 592;
 Best Local Similarity 75.0%; Pred. No. 0.036; 2; Indels 0; Gaps 0;
 Matches 12; Conservative 2; Mismatches 2;
 QY 1 PAYLRFGGKTDPLIF 16
 ||:|||||:|||||
 Db 110 PAFLRFGGKRTDPLQF 125
 RESULT 13
 Q2M1H9_HUMAN PRELIMINARY; PRT; 592 AA.
 AC Q2M1H9;
 DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
 DT 21-FEB-2006, sequence version 1.
 DT 21-FEB-2006, entry version 1.
 DE Heparanase 2.
 GN Name=HPSE2;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;
 RG NIH MGC Project;
 RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copied by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; BC112356; AAI12357.1; -; mRNA.
 SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;
 Query Match 63.5%; Score 66; DB 2; Length 592;

```
Best Local Similarity 75.0%; Pred. No. 0.036; Mismatches 2; Indels 0; Gaps 0;
Matches 12; Conservative 2;

QY 1 PAYLRFGGTKTDFLIF 16
DB 110 PAFLRFGGKRTDFLQF 125

RESULT 14
Q4TB80 TETNG
ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80; 2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG0003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Blomont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAB01007180; CAF89852.1; -; Genomic_DNA.
FT NON TER 597 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match 63.5%; Score 66; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 0.037;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIF 16
DB 105 PAFLRFGGKRTDFLQF 120

RESULT 15
Q7MH05 VIBVY
ID Q7MH05 VIBVY PRELIMINARY; PRT; 326 AA.
AC Q7MH05;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE HfIC protein.
```

```
GN OrderedLocusNames=VW3068;
OS Vibrio vulnificus (strain VJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OX Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA000037; BAC95832.1; -; Genomic_DNA.
DR Biocyc; VVUL196600:VV3068-MONOMER; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR InterPro; IPR001544; Aminotrans_IV.
DR InterPro; IPR001107; Band_7.
DR InterPro; IPR010200; HfIC_
DR PANTHER; PTHR10264; Band_7; 1.
DR Pfam; PF01145; Band_7; 1.
DR SMART; SM00244; PHB; 1.
DR TIGRFAMs; TIGR01932; hfIC; 1.
KW Complete proteome.
SQ SEQUENCE 326 AA; 36883 MW; 2FC6D5F905CB22EB CRC64;

Query Match 51.0%; Score 53; DB 2; Length 326;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDPK 19
DB 293 AYEKSFGTGKDLVLDPK 310

RESULT 16
Q8DCU6 VIBVU
ID Q8DCU6 VIBVU PRELIMINARY; PRT; 326 AA.
AC Q8DCU6;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Membrane protease subunits.
GN OrderedLocusNames=VV11297; ORFNames=VV1_1297;
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6 ";
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE016795; AA009752.1; -; Genomic_DNA.
DR Biocyc; VVUL216895:VV11297-MONOMER; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
```

DR GO: 0005515; F:protein binding; IEA.
 DR GO: 0008152; P:metabolism; IEA.
 DR InterPro: IPR001544; AminoTrans_IV.
 DR InterPro: IPR001107; Band 7.
 DR InterPro: IPR010200; HfIC.
 DR PANTHER: PTHR10264; Band 7; 1.
 DR Pfam: PF01145; Band 7; 1.
 DR SMART: SM00244; PHB; 1.
 DR TIGRFAMs: TIGR01932; HfIC; 1.
 KW Complete proteome; Protease.
 SQ SEQUENCE 326 AA; 36864 MW; 2FDA27F01BDEC2EB CRC64;

Query Match 51.0%; Score 53; DB 2; Length 326;
 Best Local Similarity 55.6%; Pred. No. 3.2;
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDPK 19
 || : ||||| : ||||
 DB 293 AYKSFSGTKSDILVDPK 310

RESULT 17
 Q72QP6 LEPIC
 ID Q72QP6 LEPIC PRELIMINARY; PRT; 1538 AA.
 AC Q72QP6;
 DT 05-JUN-2004, integrated into UniProtKB/TrEMBL.
 DT 05-JUL-2004, sequence version 1.
 DT 07-FEB-2006, entry version 10.
 DE Hypothetical protein.
 GN OrderedLocusNames=LIC12067; ORFNames=LIC 12067;
 OS Leptospira interrogans serogroup Icterohaemorrhagiae serovar
 OS copenhageni.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroseae; Leptospira.
 OX NCBI_TaxID=44275;
 RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=Flocruz LI-130;
 RX PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;
 RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
 Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,
 Marques M.V., Oliveira M.C., Menez C.F.M., Leite L.C.C., Carrer H.,
 Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H.,
 Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Gigliotti E.A.,
 Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
 Jeronimo S.M.B., Junqueira-de-Azevedo J.L.M., Kimura E.T.,
 Kuramae E.E., Lemos E.G.M., Lemos M.V.P., Marino C.L., Nunes L.R.,
 de Oliveira R.C., Pereira G.G., Reis M.S., Schrieffer A.,
 Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
 Camargo L.E.A., Kitajima J.P., Satubal J.C., Van Sluys M.A.;
 RT "Comparative genomics of two Leptospira interrogans serovars reveals
 RT novel insights into physiology and pathogenesis.";
 RL J. Bacteriol. 186:2164-2172(2004).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

DR EMBL: AE016823; AA570638.1; -: Genomic_DNA.
 DR InterPro: IPR001589; Actnin actin bd.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1538 AA; 183383 MW; 6FF9B985E45A821C CRC64;

Query Match 51.0%; Score 53; DB 2; Length 1538;
 Best Local Similarity 50.0%; Pred. No. 18;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDP 18
 || : ||||| : ||||
 DB 486 PFYFRFGDDLDDYVSFSP 503

RESULT 18
 Q8F5E5_LEPIN

ID Q8F5E5_LEPIN PRELIMINARY; PRT; 1538 AA.
 AC Q8F5E5;
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
 DT 01-MAR-2003, sequence version 1.
 DT 07-MAR-2006, entry version 14.
 DE Hypothetical protein.
 GN OrderedLocusNames=LAI1737; ORFNames=LA_1737;
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiroseae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
 RX MEDLINE=22598143; PubMed=1212204; DOI=10.1038/nature01597;
 RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
 Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
 Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
 Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
 Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
 Saint Girons I., Somerville R.L., Wen Y.-W., Shi M.-H., Chen Z.,
 Xu J.-G., Zhao G.-P.;
 RT "Unique physiological and pathogenic features of Leptospira
 RT interrogans revealed by whole-genome sequencing.";
 RL Nature 422:888-893(2003).
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

DR EMBL: AE010300; AAN48936.1; -: Genomic_DNA.
 DR GenomeReviews: AE010300.GR; LA1737.
 DR BioCyc: LINT189518:LA1737-MONOMER; -.
 DR InterPro: IPR001589; Actnin actin bd.
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1538 AA; 183389 MW; 5E3A7868F815FA0F CRC64;

Query Match 51.0%; Score 53; DB 2; Length 1538;
 Best Local Similarity 50.0%; Pred. No. 18;
 Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDP 18
 || : ||||| : ||||
 DB 486 PFYFRFGDDLDDYVSFSP 503

RESULT 19
 Q8LHX2 ORISA
 ID Q8LHX2 ORISA PRELIMINARY; PRT; 512 AA.
 AC Q8LHX2;
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
 DT 01-OCT-2002, sequence version 1.
 DT 07-FEB-2006, entry version 9.
 DE Hypothetical protein P0022B05.130 (Hypothetical protein
 DE OSUNBA0057M23.102).
 GN Name=P0022B05.130; Synonyms=OSUNBA0057M23.102;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
 OC Ehrhartoideae; Oryzoideae; Oryzae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Sasaki T., Matsumoto T., Katayose Y.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
 RT clone:OSUNBA0057M23.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License

```
CC -----
DR EMBL; AP004262; BAC10824.1; -; Genomic_DNA.
DR EMBL; AP005165; BAD30945.1; -; Genomic_DNA.
DR Gramene; Q8LHX2; -.
DR InterPro; IPR013181; U.
KW Hypothetical protein.
SQ SEQUENCE 512 AA; 58686 MW; 6E4C74892192BD43 CRC64;

Query Match 48.1%; Score 50; DB 2; Length 512;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 YLRFGGKTDFLI FDP 18
   : ||||| : |||
Db 146 FVFEGGTPHQVLFDP 161

RESULT 20
Q4HLX4 CAMLA
ID Q4HLX4_CAMLA PRELIMINARY; PRT; 224 AA.
AC Q4HLX4;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 21-FEB-2006, entry version 6.
DE Response regulator.
OS ORFNames=CLA0531;
GN Campylobacter lari RM2100.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=306263;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RM2100;
RA Fouts D.E.; Mongodin E.F.; Mandrell R.E.; Miller W.G.; Rasko D.A.;
RA Jacques R.J.; Brinkac L.M.; DeBoy R.T.; Parker C.T.; Daugherty S.C.;
RA Dodson R.J.; Durkin A.S.; Madupu R.R.; Sullivan S.A.; Shetty J.U.;
RA Ayodeji M.A.; Shvartsbeyn A.A.; Schatz M.C.; Badger J.H.; Fraser C.M.;
RA Nelson K.E.;
RT "Major structural and novel potential virulence mechanisms from the
RT genomes of multiple Campylobacter species.";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SUBCELLULAR LOCATION: Cytoplasm (By similarity).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAFK0100002; EAL55101.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007600; P:sensory perception; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. .; IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00072; Response_reg_1.
DR Pfam; PF00486; Trans_reg_C_1.
DR ProDom; PD0000039; Response_reg_1.
DR ProDom; PD0000329; Trans_reg_C_1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS01110; RESPONSE_REGULATORY; 1.
KW Activator; DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription; Transcription regulation;
KW Two-component regulatory system.
SQ SEQUENCE 224 AA; 25674 MW; 0CDCB0C28FE6D152 CRC64;

Query Match 47.6%; Score 49.5; DB 2; Length 224;
Best Local Similarity 60.0%; Pred. No. 8.4;
Matches 12; Conservative 1; Mismatches 4; Indels 3; Gaps 1;
```

```
QY 2 AYLRFGGT---KTDFLI FDP 18
   : ||||| : |||
Db 114 ARLRFGGTNNIKIDDLVIDP 133

RESULT 21
Q6Z4S8 ORYZA
ID Q6Z4S8_ORYZA PRELIMINARY; PRT; 513 AA.
AC Q6Z4S8;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein OSJNBa0057M23.122.
GN Name=OSJNBa0057M23.122;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T.; Matsumoto T.; Katayose Y.;
RT "Oryza sativa nipponbare (GR3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBa0057M23.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP005165; BAC83785.1; -; Genomic_DNA.
DR Gramene; Q6Z4S8; -.
DR InterPro; IPR013181; U.
KW Hypothetical protein.
SQ SEQUENCE 513 AA; 59130 MW; D94C59E999A036EFD CRC64;

Query Match 47.1%; Score 49; DB 2; Length 513;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGKTDFLI FDP 18
   : ||||| : |||
Db 147 LEFGGTPRRYLFDP 161

RESULT 22
O67183 AQUAE
ID O67183_AQUAE PRELIMINARY; PRT; 588 AA.
AC O67183;
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.
DT 01-AUG-1998, sequence version 1.
DT 07-FEB-2006, entry version 22.
DE Mannosyltransferase A.
GN Name=mtfA; OrderedLocustNames=AQ_1096;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G.; Warren P.V.; Gaasterland T.; Young W.G.; Lenox A.L.;
RA Graham D.E.; Overbeek R.; Sneed M.A.; Keller M.; Aujay M.; Huber R.;
RA Feldman R.A.; Short J.M.; Olsen G.J.; Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE000657; AAC07142.1; -; Genomic_DNA.
DR PIR; D70394; D70394.
```



```
DR BiOCyc; RAE063363:AO_1096-MONOMER; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycosyltransferase; Transferase.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 588 AA; 69408 MW; E47365CA903894EF CRC64;

Query Match 47.1%; Score 49; DB 2; Length 588;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 307 YLKFKKTKTVFLVD 321

RESULT 23
Q8PS30 METWA PRELIMINARY; PRT; 161 AA.
AC Q8PS30;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Conserved protein.
GN ORFNames=MA_3257;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Goei / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AE008384; AAM32953.1; -; Genomic_DNA.
DR BiOCyc; MAZ2192952:MM3257-MONOMER; -.
KW Complete proteome.
SQ SEQUENCE 161 AA; 18820 MW; 23846110B3204F0B CRC64;

Query Match 46.2%; Score 48; DB 2; Length 161;
Best Local Similarity 52.9%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 97 PKDARFLGTPVDFIVFD 113

RESULT 24
Q8TKK5 METAC PRELIMINARY; PRT; 176 AA.
AC Q8TKK5;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Hypothetical protein.
GN ORFNames=MA_3400;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;

DR BiOCyc; RAE063363:AO_1096-MONOMER; -.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00534; Glycosyltransferase; Transferase.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 588 AA; 69408 MW; E47365CA903894EF CRC64;

Query Match 47.1%; Score 49; DB 2; Length 588;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 307 YLKFKKTKTVFLVD 321

RESULT 23
Q8PS30 METWA PRELIMINARY; PRT; 161 AA.
AC Q8PS30;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Conserved protein.
GN ORFNames=MA_3257;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Goei / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AE008384; AAM32953.1; -; Genomic_DNA.
DR BiOCyc; MAZ2192952:MM3257-MONOMER; -.
KW Complete proteome.
SQ SEQUENCE 161 AA; 18820 MW; 23846110B3204F0B CRC64;

Query Match 46.2%; Score 48; DB 2; Length 161;
Best Local Similarity 52.9%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 97 PKDARFLGTPVDFIVFD 113

RESULT 24
Q8TKK5 METAC PRELIMINARY; PRT; 176 AA.
AC Q8TKK5;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Hypothetical protein.
GN ORFNames=MA_3400;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
```

```
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferty J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AE010299; AAM06767.1; -; Genomic_DNA.
DR GenomeReviews; AE010299 GR; MA3400.
DR BiOCyc; MAE188937:MA3400-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 176 AA; 20594 MW; 19576D9B075E62C8 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 12;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 112 PKDARFLGTPVDFIVFD 128

RESULT 25
Q2JIZ0 9CYAN PRELIMINARY; PRT; 194 AA.
ID Q2JIZ0 9CYAN PRELIMINARY; PRT; 194 AA.
AC Q2JIZ0;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DE Hypothetical protein.
GN ORFNames=CVB_2470;
OS Cyanobacteria bacterium Yellowstone B-Prime.
OC Bacteria; Cyanobacteria.
OX NCBI_TaxID=331115;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Yellowstone B-Prime;
RA Heidelberg J.;
RL Submitted (DEC-2005) to the EMBL/GenBank/DBDJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: CP000240; ABD03404.1; -; Genomic_DNA.
DR Hypothetical protein.
SQ SEQUENCE 194 AA; 22088 MW; 0DA56BDEA16427D CRC64;

Query Match 46.2%; Score 48; DB 2; Length 194;
Best Local Similarity 58.8%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFD 17
||:| ||||| ||:|
DB 117 POQLAFQKTKGDFLIFE 133

RESULT 26
Q86YUO_HUMAN
```

```

ID Q86YU0_HUMAN PRELIMINARY; PRT; 208 AA.
AC Q86YU0;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Carbonic anhydrase VII short form.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP Chen Y., Huang C.-H.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AY075020; AAL78168.1; -; mRNA.
DR HSPSP; P00918; 1UGD.
DR Ensembl; ENSG00000168748; Homo sapiens.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_Coanhd.
DR PANTHER; PTHR18952; Euk_Coanhd.
DR Pfam; PF00194; Carb anhydrase; 1.
DR ProDom; PD000865; Euk_Coanhd; 1.
DR PROSITE; PS00162; ALPHA CA 1; 1.
DR SEQUENCE 208 AA; 23452 MW; AF3D016A27182D18 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 208;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLIFDPK 19
   :||| ||| | | | | |
Db 113 VRFKGTAKQFSCFNPK 128

RESULT 27
ID Q2RP20_RHURU PRELIMINARY; PRT; 247 AA.
AC Q2RP20;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Cytochrome c oxidase, mono-heme subunit/FixO.
GN ORFNames=Rru_A3331;
OS Rhodospirillum rubrum ATCC 11170.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Rhodospirillum.
OX NCBI_TaxID=269796;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 11170;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Munk A.C., Brettin T., Bruce D.,
RA Han C., Tapia R., Gilna P., Schmutz J., Larimer F., Land M.,
RA Kyrpides N., Mavromatis K., Richardson P., Zhang Y., Roberts G.,
RA Reslewic S., Zhou S., Schwartz D.C.;
RT "Complete sequence of the chromosome of Rhodospirillum rubrum ATCC
RT 11170."
RL Submitted (DEC-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000230; ABC24125.1; -; Genomic DNA.
DR SEQUENCE 247 AA; 27561 MW; 359BFD038B0483A1 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 247;

```

```

Best Local Similarity 58.8%; Pred. No. 17;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTKDFLIFDP 18
   |||: ||| ||| |||
Db 224 AYQLVLGTMVDFQTQDP 240

RESULT 28
CAH7_HUMAN
ID CAH7_HUMAN STANDARD; PRT; 264 AA.
AC P43166; Q541F0;
DT 01-NOV-1995, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1995, sequence version 1.
DT 07-FEB-2006, entry version 44.
DE Carbonic anhydrase 7 (EC 4.2.1.1) (Carbonic anhydrase VII) (Carbonate
DE dehydratase VII) (CA-VII).
GN Name=CA7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=92147127; PubMed=1783392;
RA Montgomery J.C., Venta P.J., Eddy R.L., Fukushima Y.S., Shows T.B.,
RA Tashian R.E.;
RT "Characterization of the human gene for a newly discovered carbonic
RT anhydrase, CA VII, and its localization to chromosome 16."
RL Genomics 11:835-848(1991).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA].
RA Chen Y., Huang C.-H.;
RT "Molecular identification of carbonic anhydrases (CA) and CA-related
RT (CAK) genes."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=99425270; PubMed=10493829; DOI=10.1006/geno.1999.5927;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q."
RL Genomics 60:295-308(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Colon, Kidney, and Stomach;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussdin T.B., Ioshizuka S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Dickson M.C.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Reversible hydration of carbon dioxide.
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.

```

```
CC CC -|- COPACTOR: Zinc (By similarity).
CC CC -|- SUBCELLULAR LOCATION: Cytoplasm (Probable).
CC CC -|- SIMILARITY: Belongs to the alpha-carbonic anhydrase family.
CC CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: M76423; AAAS1923.1; -; Genomic DNA.
CC DR EMBL: M76420; AAAS1923.1; JOINED; Genomic DNA.
CC DR EMBL: M76421; AAAS1923.1; JOINED; Genomic DNA.
CC DR EMBL: M76422; AAAS1923.1; JOINED; Genomic DNA.
CC DR EMBL: AY075019; AAL78167.1; -; mRNA.
CC DR EMBL: AC004638; AAC23785.1; -; Genomic DNA.
CC DR EMBL: BC033865; AAH33865.1; -; mRNA.
CC DR PIR: A55272; CRU07.
CC DR HSSP: P00918; 1BV3.
CC DR Ensembl: ENSG00000168748; Homo sapiens.
CC DR HGNC: HGNC:1381; CAV.
CC DR MIM: 114770; Gene.
CC DR GO: GO:0004089; F:carbonate dehydratase activity; TAS.
CC DR InterPro: IPR001148; Euk Coanhd.
CC DR PANTHER: PTHR18952; Euk Coanhd; 1.
CC DR Pfam: PF00194; Carb. anhydrase; 1.
CC DR ProDom: PD000865; Euk Coanhd; 1.
CC DR PROSITE: PS00162; ALPHA_CA_1; 1.
CC DR PROSITE: PS1144; ALPHA_CA_2; 1.
CC DR Lyase; Metal-binding; Zinc.
KW CHAIN 1 264
FT FT Carbonic anhydrase 7.
FT FT /FTID=PRO_0000077431.
FT METAL 96 96 Zinc (catalytic) (By similarity).
FT METAL 98 98 Zinc (catalytic) (By similarity).
FT METAL 121 121 Zinc (catalytic) (By similarity).
SQ SEQUENCE 284 AA; 29658 MW; 7AD559FC6E07EF96 CRC64;

Query Match 46.2%; Score 48; DB 1; Length 264;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 LRFPGTKYDFLIIDPK 19
Db 169 VRFKGTAKQFSCFNPK 184

RESULT 29
Q928Y1 CHLPN PRELIMINARY; PRT; 265 AA.
AC Q928Y1; 07AJ65; Q7BXP9; Q7DEH6;
DT 01-MAY-1999, integrated into UniProtKB/TrEMBL.
DT 01-MAY-1999, sequence version 1.
DT 21-FEB-2006, entry version 30.
DE Hypothetical protein CP0564 (Hypothetical protein CPJ0203).
GN OrderedLocustNames=CP0564, CPJ0203, CPB0207; ORFNames=CPn_0203;
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388; DOI=10.1038/7716;
RA Kaiman S., Mitchell W.P., Marathe R., Lemmel C.J., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389 (1999).
RN [2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935; DOI=10.1093/nar/28.6.1397;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J.F., Uterback T.R., Berry K.J.,
RA Baas S., Linher K.D., Weidman J.F., Khouri H.M., Craven B., Bowman C.,
RA Dodson K.J., Gwin M.L., Nelson W.C., DeBoy R.T., Kolonay J.F.,
RA McClarty G., Salzberg S.L., Eisen J.A., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```

```
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=TW-183;
RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362; DOI=10.1093/nar/28.12.2311;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314 (2000).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AE001363; AAD18356.1; -; Genomic DNA.
CC EMBL: AE002215; AAF38383.1; -; Genomic DNA.
CC EMBL: AE017157; AAP98140.1; -; Genomic DNA.
CC EMBL: BA000008; BAA98413.1; -; Genomic DNA.
CC DR PIR: C86516; C86516.
CC DR PIR: D72105; D72105.
CC DR TIGR: CP0564; -.
CC DR BioCyc: CPNE115711:CP0564-MONOME-; -.
CC DR BioCyc: CPNE115713:CPN0203-MONOM-; -.
CC DR BioCyc: CPNE138677:CPN0203-MONOM-; -.
CC DR BioCyc: CPNE182082:CPB0207-MONOM-; -.
CC DR InterPro: IPR006974; DUF648.
CC DR Pfam: PF04890; DUF648; 1.
CC Complete proteome; Hypothetical protein.
KW SEQUENCE 265 AA; 30482 MW; 36B4C6B4055267A1 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 265;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 AYLRFGTKYDFLIIDFP 18
Db 24 SYFFFGTTRTQILVITP 40

RESULT 30
Q41BD7_9BACI PRELIMINARY; PRT; 311 AA.
AC Q41BD7;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE PfkB.
GN ORFNames=ExigDRAFT_0369;
OS Exiguobacterium sp. 255-15.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Exiguobacterium.
OX NCBI_TaxID=262543;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Exiguobacterium sp.
RT 255-15.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-ORNL);
```

```
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Exiguobacterium sp. 255-15";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T., Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AADW02000018; EAM86641.1; -; Genomic_DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR002173; PfKB.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
KW Kinase; Transferase.
SQ SEQUENCE 311 AA; 33430 MW; 37F2CA05DD010579 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 311;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 GTKTDLFILFDPK 19
DB 68 GVKTDHLVDFDPE 79

RESULT 31
ID Q8G431_BIFLO PRELIMINARY; PRT; 411 AA.
AC Q8G431;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Probable aminotransferase.
GN OrderedLocustNames=BL1564;
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787; DOI=10.1073/pnas.212527599;
RA Schell M.A., Karimantou M., Snel B., Vilanova D., Berger B., Passi G., Zahren M.-C., Desiere F., Bork P., Delley M., Pridmore R.D., Arigoni F.;
RL "The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AE014295; AAN25355.1; -; Genomic_DNA.
DR BiOCyc; BLON206672:BL1564-MONOMER; -.
DR GO; GO:0008483; F:transaminase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
```

```
DR InterPro; IPR004839; Aminotrans_I/II.
DR Pfam; PF00155; Aminotran_1_2; 1.
KW Aminotransferase; Complete proteome; Transferase.
SQ SEQUENCE 411 AA; 43070 MW; EC2F7E468523A26D CRC64;

Query Match 46.2%; Score 48; DB 2; Length 411;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTD 12
DB 388 PAYLRFSAFATD 399

RESULT 32
Q3FBC5_9BURK PRELIMINARY; PRT; 425 AA.
AC Q3FBC5;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Mandelate racemase/muconate lactonizing enzyme.
GN ORFNames=BambDRAFT_3416;
OS Burkholderia ambifaria AMMD.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=339670;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T., Hammon N., Israni S., Pitluck S., Richardson P.;
RL "Sequencing of the draft genome and assembly of Burkholderia ambifaria AMMD.";
RT AMMD.;
RN Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia ambifaria AMMD.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAJ01000005; EAO46342.1; -; Genomic_DNA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001354; MR_MLE.
DR Pfam; PF01188; MR_MLE; 1.
SQ SEQUENCE 425 AA; 45842 MW; CC9E0BB8836F7C32 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 425;
Best Local Similarity 56.2%; Pred. No. 31;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 5 RFGGTTK--DFLIDPP 18
DB 313 RYGGURADRDFLVDFP 328

RESULT 33
Q8T108_BOMMO PRELIMINARY; PRT; 515 AA.
ID Q8T108;
AC Q8T108;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
```

```
DT 07-FEB-2006, entry version 10.
DE Heparanase-like protein.
GN Name=Enhepa;
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=sp50; TISSUE=Posterior silkland;
RA Koike Y., Mita K., Suzuki M.G., Maeda S., Abe H., Osoegawa K.,
RA deJong P.J., Shimada T.;
RT "Genomic sequence of a 320-kb segment of the Z chromosome of Bombyx
RT mori containing a kettin ortholog.";
RL Mol. Genet. Genomics 269:137-149(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB079860; BAB95191.1; -; Genomic DNA.
DR EMBL; AB090307; BAC10612.1; -; Genomic DNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 515 AA; 59770 MW; FB8100ABE6EDDADB CRC64;

Query Match 46.2%; Score 48; DB 2; Length 515;
Best Local Similarity 62.5%; Pred. No. 38;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDLFIF 16
Db 84 PARLRLGGTMSERLIF 99

RESULT 34
Q47YDS_COLP3
ID Q47YDS_COLP3 PRELIMINARY; PRT; 638 AA.
AC Q47YDS;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Prolyl oligopeptidase family protein (EC 3.4.21.26).
GN OrderedLocusNames=CPS_3511;
OS Colwellia psychrerythraea (strain 34H / ATCC BAA-681) (Vibrio
OS psychrerythrus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Colwelliaceae; Colwellia.
OX NCBI_TaxID=167879;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043709; DOI=10.1073/pnas.0504766102;
RA Mehta B.A., Nelson K.E., Deming J.W., Momen B., Melamud E., Zhang X.,
RA Mout R.J., Madupu R., Nelson W.C., Dodson R.J., Brinkac L.M.,
RA Dougherty S.C., Durkin A.S., DeBoy R.T., Kolonay J.F., Sullivan S.A.,
RA Zhou L., Daviden T.M., Wu M., Huston A.L., Lewis M., Weaver B.,
RA Weidman J.F., Khouri H., Uterback T.R., Feldblyum T.V., Fraser C.M.;
RT "The psychrophilic lifestyle as revealed by the genome sequence of
RT Colwellia psychrerythraea 34H through genomic and proteomic
RT analyses.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:10913-10918(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000083; AAZ24771.1; -; Genomic DNA.
DR TIGR; CPS_3511; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004287; F:prolyl oligopeptidase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR00379; Ser esters.
DR Pfam; PF00326; Peptidase_S9; 1.

KW Complete proteome; Hydrolase.
SQ SEQUENCE 638 AA; 71717 MW; 59DDA8B6C3BE4821 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 638;
Best Local Similarity 57.1%; Pred. No. 49;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 LRGGTKTDLFIFD 17
Db 598 IRSGGTVDYLVFD 611

RESULT 35
Q8KZS8_ACEPA
ID Q8KZS8_ACEPA PRELIMINARY; PRT; 742 AA.
AC Q8KZS8;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Alcohol dehydrogenase.
OS Acetobacter pasteurianus (Acetobacter turbidans).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acetobacter.
OX NCBI_TaxID=438;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCI 1193;
RA Takakuwa N., Yamane K., Oda Y., Fukaya M., Tsukamoto Y., Ohnishi M.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB086012; BAB97167.1; -; Genomic DNA.
DR HSP; Q46444; 1KB0.
DR GO; GO:0020288; C:periplasmic space (sensu Gram-negative Bact. .; IEA.
DR GO; GO:0020037; F:heme binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR009056; Cyt c monohaem.
DR InterPro; IPR012282; Cytochrome_c_R.
DR InterPro; IPR001479; PQQ_bac.
DR InterPro; IPR002372; PQQ_repeat.
DR Pfam; PF01011; PQQ; 3.
DR SMART; SM00564; PQQ; 2.
DR PROSITE; PS00363; BACTERIAL_PQQ_1; 1.
DR PROSITE; PS51007; CYTC; 1.
SQ SEQUENCE 742 AA; 81802 MW; 42A7A50DCE2D981 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 742;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GGTGKTDLFIFDP 18
Db 276 GGTGVDLFVYDP 287

RESULT 36
Q2IRUO_RHOPA
ID Q2IRUO_RHOPA PRELIMINARY; PRT; 838 AA.
AC Q2IRU0;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Periplasmic sensor signal transduction histidine kinase precursor.
GN ORFNames=RPB_4383;
OS Rhodopseudomonas palustris Haa2.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodospseudomonas.
OX NCBI_TaxID=316058;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Haa2;
```

```
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Hauser L., Pelletier D.A.,
RA Kyrpides N., Anderson I., Oda Y., Harwood C.S., Richardson P.;
RT "Complete sequence of Rhodopseudomonas palustris Ha22.";
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; CP000250; ABD09070.1; --; Genomic_DNA.
KW Kinase; Signal.
FT SIGNAL 1 25 Potential.
SQ SEQUENCE 838 AA; 90636 MW; 0D237021D5ECB306 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 838;
Best Local Similarity 64.3%; Pred. No. 66;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 FGGTGTGRFTIFDPK 19
Db 315 FGGTGTGRFTIFDPR 328

RESULT 37
Q3QRL3_9RHOB PRELIMINARY; PRT; 1895 AA.
ID Q3QRL3_9RHOB PRELIMINARY; PRT; 1895 AA.
AC Q3QRL3;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Outer membrane autotransporter barrel.
GN ORFNames=RoseDRAFT_1391;
OS Silicibacter sp. TW1040.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Silicibacter.
OX NCBI_TaxID=292414;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Silicibacter sp.
TW1040.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Silicibacter sp. TW1040.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAFG02000006; EAN5260.1; --; Genomic_DNA.
DR InterPro; IPR006034; Asp/Glutamase.
DR InterPro; IPR005546; Auto transptbeta.
DR InterPro; IPR006315; Autotransporter.
DR InterPro; IPR001298; Filamin.
```

```
DR InterPro; IPR001969; Pept_Asp_AS.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRfams; TIGR01414; autotrans_bar1; 1.
DR PROSITE; PS00144; ASN_GLN_ASE_1; UNKNOWN 3.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS00194; FILAMIN_REPEAT; 2.
SQ SEQUENCE 1895 AA; 188024 MW; 2A2C185A0FBD7866 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 1895;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 GGKTKDFLIFDP 18
Db 101 GGSSTNFFIFDP 112

RESULT 38
Q2LZG0_DROPS PRELIMINARY; PRT; 2405 AA.
ID Q2LZG0_DROPS PRELIMINARY; PRT; 2405 AA.
AC Q2LZG0;
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 1.
DE GA19954-PA (Fragment).
GN Name=Dpse\GA19954; ORFNames=Dpse GA19954;
OS Drosophila pseudoobscura (Fruit Fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RX PubMed=15632085; DOI=10.1101/gr.3059305;
RA Richards S., Liu Y., Bettencourt B.R., Hradecky P., Letovsky S.,
RA Nielsen R., Thornton K., Hubisz M.J., Chen R., Meisel R.P.,
RA Couronne O., Hua S., Smith M.A., Zhang P., Liu J., Bussemaker H.J.,
RA van Batenburg M.F., Howells S.L., Scherer S.E., Sodergren E.,
RA Matthews B.B., Crosby M.A., Schroeder A.J., Ortiz-Barrientos D.,
RA Rives C.M., Metzker M.L., Muzny D.M., Scott G., Steffen D.,
RA Wheeler D.A., Worley K.C., Havlak P., Durbin K.J., Egan A., Gill R.,
RA Hume J., Morgan M.B., Miner G., Hamilton C., Huang Y., Waldron L.,
RA Verduzco D., Clerc-Blankenburg K.P., Dubchak I., Noor M.A.F.,
RA Anderson W., White K.P., Clark A.G., Schaeffer S.W., Gelbart W.,
RA Weinstock G.M., Gibbs R.A.;
RT "Comparative genome sequencing of Drosophila pseudoobscura:
chromosomal, gene, and cis-element evolution.";
RL Genome Res. 15:1-18(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RG FlyBase;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RG Human Genome Sequencing Center;
RA Richards S., Liu Y., Bettencourt B.R., Hradecky P., Letovsky S.,
RA Chen R., Smith M.A., Howells S.L., Scherer S.E., Sodergren E.,
RA Rives C.M., Metzker M.L., Munzy D.M., Wheeler D.A., Worley K.C.,
RA Havlak P., Durbin K.J., Egan A., Gill R., Hume J., Morgan M.B.,
RA Huang Y., Waldron L., Verduzco D., Blankenburg K.P., Adams C.,
RA Allen C., Allen H., Anyalebechi V., Asomugha C., Bellard T.,
RA Bhuchar V., Biswal K., Blair J., Blomstrom D., Burrell K.,
RA Calderon E., Cardenas V., Carter K., Cavazos I., Ceasar H.,
RA Chavez D., Chu J., Cockrell R., Cox C., Coyle M., Davila M., Davis C.,
RA Davy-Carroll L., De A., Delgado O., Denson S., Deramo C., Dinh H.,
RA Eaves K., Escotto M., Eugene C., Falls T., Fernandez S., Flagg N.,
RA Forbes L., Garner T., Garza M., Ghose S., Grady M., Hamilton C.,
RA Hernandez J., Hines S., Hogues M., Hollins B., Idlebird D., Iino K.,
RA Jimenez A., Johnson B., Jolivet A., Kelly S., King L., Kisano H.,
RA Kovar C., Lebow H., Lee K., LeGall F., Lewis L., Li Z., London P.,
```

```
RA Lopez J., Lozado R., Malloy K., Martinez E., Mercadao C., Miner G.,
RA Minja E., Moore S., Navavati A., Ngo R., Nguyen N., Nwaokelenah O.,
RA Okwundu G., Parks K., Pasternak S., Patel B., Paul H., Payne C.,
RA Poindexter A., Primus E., Pu L.-L., Puazo M., Quiroz J., Rabata D.,
RA Reigh R., Ruiz S., Sanders W., Sison I., Sorelle R., Taylor C.,
RA Taylor T., Thomas N., Trejos Z., Usmani K., Vera V., Villasana D.,
RA Wang S., Warren J., Warren R., White P., Wlezyk R., Wright R.,
RA Noor M.A.F., Schaeffer S.W., Gelbart W., Weinstein G.M., Gibbs R.A.,
RA Weinstein G., Gibbs R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CH379069; EAL29548.1; -; Genomic_DNA.
FT NON TER 1
FT TER 2405 2405
SQ SEQUENCE 2405 AA; 256997 MW; 0456101FDB2A668B CRC64;

Query Match 46.2%; Score 48; DB 2; Length 2405;
Best Local Similarity 52.6%; Pred. No. 2.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDFLIFDPK 19
Db 2138 PASLYFSQTDRLDLSBPK 2156

RESULT 39
Q9WZB4.THEMA PRELIMINARY; PRT; 176 AA.
ID Q9WZB4.THEMA
AC Q9WZB4;
DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1999, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Hypothetical protein.
GN OrderedLocNames=TW0646; ORFNames=TW_0646;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
RA Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
RN Nature 399:323-329(1999).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB000512; AAD35730.1; -; Genomic_DNA.
DR PIR; E72351; E72351.
DR TIGR; TM0646; -.
DR BioCyc; TWAR2336:TM0646-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 176 AA; 2195 MW; D5768886C85A08FB CRC64;

Query Match 45.2%; Score 47; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDFLIFD 17
Db 111 PKDARFIGTPVDFVVD 127

us-10-645-659a-7.rup
```

```
RESULT 40
Q5ZIQ7.CHICK PRELIMINARY; PRT; 1076 AA.
ID Q5ZIQ7.CHICK
AC Q5ZIQ7;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Hypothetical protein.
GN ORFNames=RCUMB04_24e3;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT genefunction analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ720727; CAG32386.1; -; mRNA.
DR GO; GO:0005488; F:binding; IEA.
DR InterPro; IPR011989; ARM-like.
DR InterPro; IPR005612; CBF.
DR Pfam; PF03914; CBF; 1.
KW Hypothetical protein.
SQ SEQUENCE 1076 AA; 122590 MW; 683AE5974F95A164 CRC64;

Query Match 45.2%; Score 47; DB 2; Length 1076;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRFGGTKTDFLI 15
Db 45 LRLGGTKQDFLM 56

RESULT 41
Q5ZMV4.CHICK PRELIMINARY; PRT; 1076 AA.
ID Q5ZMV4.CHICK
AC Q5ZMV4;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Hypothetical protein.
GN ORFNames=RCUMB04_1b8;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatski A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT genefunction analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ719280; CAG30939.1; -; mRNA.
DR GO; GO:0005488; F:binding; IEA.
```

```
DR InterPro; IPR011989; ARM-like.
DR InterPro; IPR005612; CBF.
DR Pfam; PF03914; CBF; 1.
KW Hypothetical protein.
SQ SEQUENCE 1076 AA; 122549 MW; 7752F6970B8DD267 CRC64;

Query Match 45.2%; Score 47; DB 2; Length 1076;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLI 15
Db 45 LRLGGTKQDFLM 56

RESULT 42
Q33FF2 METHU PRELIMINARY; PRT; 180 AA.
AC Q33FF2;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Conserved protein.
GN ORFNames=MhndRAFT_0483;
OS Methanospirillum hungatei JF-1.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales;
OC Methanospirillaceae; Methanospirillum.
OX NCBI_TaxID=323259;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JF-1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Methanospirillum
RT hungatei JF-1.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JF-1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Methanospirillum
RT JF-1.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AALU01000044; EAP14488.1; -; Genomic DNA.
DR EMBL; AALU01000044; EAP14488.1; -; Genomic DNA.
SQ SEQUENCE 180 AA; 21279 MW; 260C336E54315975 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 180;
Best Local Similarity 47.1%; Pred. No. 26;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLI 17
Db 118 PSDARFLGSPVDIFVD 134

RESULT 43
Q4CAE2 CROWT PRELIMINARY; PRT; 219 AA.
ID Q4CAE2_CROWT PRELIMINARY; PRT; 219 AA.
AC Q4CAE2;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE Lipote-protein ligase B.
GN ORFNames=CwatDRAFT_5612;
```

```
OS Crocosphaera watsonii.
OC Bacteria; Cyanobacteria; Chroococcales; Crocosphaera.
OX NCBI_TaxID=165597;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Crocosphaera watsonii
RT WH 8501.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Crocosphaera watsonii WH
RT 8501.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AADV02000001; EAM52994.1; -; Genomic DNA.
DR GO; GO:0008415; F:acetyltransferase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009107; P:lipoate biosynthesis; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR000544; Lipoteate_B.
DR Pfam; PF03099; BFL_LipA_LipB; 1.
DR ProDom; PD006086; Lipoteate_B; 1.
DR TIGRFAMs; TIGR00214; lipB; 1.
DR PROSITE; PS01313; LipB; 1.
KW Ligase.
SQ SEQUENCE 219 AA; 25080 MW; F44134A46CF65B9D CRC64;

Query Match 44.2%; Score 46; DB 2; Length 219;
Best Local Similarity 47.4%; Pred. No. 33;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLI 19
Db 53 PVTYLTGSGSTKFKIFDPK 71

RESULT 44
Q5TMQ4 ANOGA PRELIMINARY; PRT; 308 AA.
ID Q5TMQ4_ANOGA PRELIMINARY; PRT; 308 AA.
AC Q5TMQ4;
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE ENSANGP00000026025.
GN ORFNames=ENSANGG00000022402;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1];
RP NUCLEOTIDE SEQUENCE.
```



```

RC STRAIN=PEST;
RG The Anopheles gambiae re-annotation.;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAAB01008986; EAL38841.1; -; Genomic DNA.
DR GO; GO:000175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001247; ExoRNase.
DR Pfam; PF01138; RNase PH; 1.
DR SEQUENCE 308 AA; 33784 MW; 3A2F1F2B872F8FA7 CRC64;
SQ
Query Match 44.2%; Score 46; DB 2; Length 308;
Best Local Similarity 44.4%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDPK 19
DB ||: |||: |||:
69 AYLGLGNTKIVSVFDPK 86

RESULT 45
Q2W843_MAGSA
ID Q2W843_MAGSA PRELIMINARY; PRT; 319 AA.
AC Q2W843;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DE 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=amb1178;
OS Magnetospirillum magneticum AMB-1.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Magnetospirillum.
OX NCBI_TaxID=342108;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMB-1;
RX PubMed=16303747;
RA Matsunaga T., Okamura Y., Fukuda Y., Wahyudi A.T., Murase Y.,
RA Takeyama H.;
RA "Complete Genome Sequence of the Facultative Anaerobic Magnetotactic
RT Bacterium Magnetospirillum sp. strain AMB-1.";
RL DNA Res. 12:157-166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AP007255; BA849982.1; -; Genomic_DNA.
DR Hypothetical protein.
KW
DR SEQUENCE 319 AA; 33449 MW; FDA8F64758588E5D CRC64;
SQ
Query Match 44.2%; Score 46; DB 2; Length 319;
Best Local Similarity 52.9%; Pred. No. 50;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 PAYLRFGGKTDFLIFD 17
DB ||: |||: |||:
126 PKFRRHGAKLKDLSFD 142

RESULT 46

```

```
Db      162 FIRFGGKRTTALYFEP 177
      ::|||::|::|::|
RESULT 48
QSPDY6_SALPA PRELIMINARY; PRT; 404 AA.
ID QSPDY6
AC QSPDY6;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocusNames=SPA0754;
OS Salmonella paratyphi-a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=54388;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 9150 / SARB42;
RX PubMed=15531882; DOI=10.1038/ng1470;
RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,
  Porwollik S., Sabo A., Meyer R., Bieri T., Ozersky P., McLellan M.,
  Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
  Kohlberg S., Strong C., Du F., Carter J., Kremizki C., Layman D.,
  Leonard S., Sun H., Fulton L., Nash W., Miner T., Minx P.,
  Dalehaunt K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
  Spieth J., Wilson R.K.;
RA "Comparison of genome degradation in Paratyphi A and Typhi, human-
  restricted serovars of Salmonella enterica that cause typhoid.";
RL Nat. Genet. 36:1268-1274(2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; CP000026; AAU76750.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 404 AA; 45129 MW; AB77A8DCB1812351 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 YLRFGGTTKDFLIFDP 18
      ::|||::|::|::|
Db      162 FIRFGGKRTTALYFEP 177

RESULT 49
Q9F7A7_SALTY PRELIMINARY; PRT; 404 AA.
ID Q9F7A7_SALTY
AC Q9F7A7; Q7CQA2;
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2001, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Hypothetical protein wcaD (Putative colanic acid polymerase).
GN Name=wcaD; OrderedLocusNames=STM2112;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LT2;
RX PubMed=20461159; PubMed=11004393;
RA Stevenson G., Lan R., Reeves P.R.;
RT "The colanic acid gene cluster of Salmonella enterica has a complex
  history.";
RL FEMS Microbiol. Lett. 191:11-16(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=LT2 / SGCSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
```

```
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
  Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
  Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
  Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
  Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
  LT2.";
RL Nature 413:852-856(2001).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; AF285084; AAG24810.1; -; Genomic_DNA.
DR EMBL; AE008793; AAL21016.1; -; Genomic_DNA.
DR BioCyc; STYV99287:STM2112-MONOMER; -.
DR LinkHub; Q9F7A7; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 45147 MW; AA72F888E36D0D91 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 YLRFGGTTKDFLIFDP 18
      ::|||::|::|::|
Db      162 FIRFGGKRTTALYFEP 177

RESULT 50
WCAD_ECOLI STANDARD; PRT; 405 AA.
ID WCAD_ECOLI
AC P71238; P76385;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 2.
DT 07-MAR-2006, entry version 34.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocusNames=b2056;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Adrianopoulos K., Hobbs M., Reeves P.R.;
RT "Organization of the Escherichia coli K-12 gene cluster responsible
  for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [2]
RP SEQUENCE REVISION.
RC STRAIN=K12;
RA Reeves P.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
  Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SUBCELLULAR LOCATION.
RC STRAIN=K12 / MG1655;
RX PubMed=15919996; DOI=10.1126/science.1109730;
RA Daley D.O., Rapp M., Granseth E., Melen K., Drew D., von Heijne G.;
RT "Global topology analysis of the Escherichia coli inner membrane
  proteome.";
RL Science 308:1321-1323(2005).
CC -!- PATHWAY: Slime polysaccharide colanic acid biosynthesis.
```


GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 129.765 Seconds
(without alignments)
3728.138 Million cell updates/sec

Title: US-10-645-659A-5

Perfect score: 2728

Sequence: 1 MVLVLLVLLVAVPPRTAE.....LPAFSYGFYVIRNAKAIACI 523

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.*

1: uniprot_prot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2728	100.0	523	1 HPSE_CHICK	Q90YK5 gallus gall
2	1653	60.6	574	2 Q333X6_SPAJD	Q333X6 spalax juda
3	1649	60.4	574	2 Q333X7_9RODE	Q333X7 spalax carm
4	1648.5	60.4	543	1 HPSE_HUMAN	Q9Y251 homo sapien
5	1641	60.2	574	2 Q333X9_9RODE	Q333X9 spalax gali
6	1640	60.1	574	2 Q333X8_9RODE	Q333X8 spalax gola
7	1624	59.5	545	1 HPSE_BOVIN	Q9MY70 bos taurus
8	1602	58.7	535	1 HPSE_MOUSE	Q6Y9Z1 mus musculus
9	1572	57.6	536	1 HPSE_RAT	Q711P1 rattus norv
10	1558	57.1	558	2 Q333X5_SPAJD	Q333X5 spalax juda
11	1232	45.2	533	2 Q4SYF6_TETNG	Q4SYF6 tetraodon n
12	1028.5	37.7	592	1 HPSE2_HUMAN	Q8WQZ2 homo sapien
13	1028.5	37.7	592	2 Q2MIH9_HUMAN	Q2MIH9 homo sapien
14	936.5	34.3	597	2 Q4TB80_TETNG	Q4TB80 tetraodon n
15	679	24.9	255	2 Q4T8C8_TETNG	Q4T8C8 tetraodon n
16	640	23.5	515	2 Q8T108_BOMMO	Q8T108 bombyx mori
17	395.5	14.5	539	2 Q2QN56_ORYSA	Q2QN56 oryza sativ
18	395	14.5	543	1 HPSE1_ARATH	Q9FF10 arabidopsis
19	387.5	14.2	529	2 Q6ZJE2_ORYSA	Q6ZJE2 oryza sativ
20	382	14.0	526	2 Q5SNA6_ORYSA	Q5SNA6 oryza sativ
21	375.5	13.8	537	2 Q70XJ3_HORVU	Q70XJ3 hordeum vul
22	351.5	12.9	541	2 Q691I6_ORYSA	Q691I6 oryza sativ
23	350.5	12.8	536	1 HPSE3_ARATH	Q9FZP1 arabidopsis
24	346	12.7	527	2 Q9LRC8_SCUBA	Q9LRC8 scutellaria
25	339	12.4	539	1 HPSE2_ARATH	Q81608 arabidopsis
26	336.5	12.3	401	2 Q303Z4_ARATH	Q303Z4 arabidopsis
27	336.5	12.3	559	2 Q89F99_BRAJA	Q89F99 bradyrhizob
28	317.5	11.6	516	2 Q447R5_SOLUS	Q447R5 solibacter
29	289	10.6	506	2 Q37Q70_SPHAR	Q37Q70 novosphingo
30	257	9.4	382	2 Q3EBP7_ARATH	Q3EBP7 arabidopsis
31	249.5	9.1	537	2 Q43803_SOLUS	Q43803 solibacter

32	186.5	6.8	463	2	Q63T97_BURPS	Q63T97 burkholderi
33	185.5	6.8	670	2	Q3JTG0_BURP1	Q3JTG0 burkholderi
34	139.5	5.1	1128	2	Q5TT65_ANOGA	Q5TT65 anopheles g
35	134.5	4.9	510	2	Q2U0T3_ASPOR	Q2U0T3 aspergillus
36	129	4.7	935	2	Q9VE79_DROME	Q9VE79 drosophila
37	127	4.7	484	2	Q5RFE6_PONPY	Q5RFE6 pongo pygma
38	122.5	4.5	533	2	Q7N6A1_PHOLL	Q7N6A1 photorhabdu
39	120	4.4	559	2	Q7SFB0_NEUCR	Q7SFB0 neurospora
40	120	4.4	597	2	Q6ZEB8_BURMA	Q6ZEB8 burkholderi
41	120	4.4	597	2	Q62IW2_BURPS	Q62IW2 burkholderi
42	120	4.4	642	2	Q3JUN8_BURP1	Q3JUN8 burkholderi
43	119.5	4.4	533	2	Q7N666_PHOLL	Q7N666 photorhabdu
44	119	4.4	1630	2	Q5B9V9_EMENI	Q5B9V9 aspergillus
45	117.5	4.3	373	2	Q39AY5_BURS3	Q39AY5 burkholderi

ALIGNMENTS

RESULT 1

HPSE_CHICK ID HPSE_CHICK STANDARD; PRT; 523 AA.
AC Q90YK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP MEDLINE=2136959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldshmidt O., Zoharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Becker I., Mitrani E., Vlodavsky I.;
RT "Expression pattern and secretion of human and chicken heparanase are
RT determined by their signal peptide sequence.";
J. Biol. Chem. 276:29178-29187(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC (By similarity).
CC -!- PTM: N-glycosylated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NonCommercial License

EMBL; AY037007; AA82648.1; -, mRNA.
Ensembl; ENSGALG00000011203; Gallus gallus.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 523 Heparanase.
FT REGION 137 141 /FTid=PRO_0000042259.
FT REGION 250 260 Heparin/HS-binding (By similarity).
FT ACT_SITE 204 204 Heparin/HS-binding (By similarity).
FT ACT_SITE 323 323 Proton donor (Potential).
FT CARBOHYD 141 141 Nucleophile (Potential).
FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 100.0%; Score 2728; DB 1; Length 523;
Best Local Similarity 100.0%; Pred. No. 5.6e-207;

```
Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLVLLVLLVLLVPPRTTAEQLGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKL 60
DB 1 MLVLLVLLVLLVPPRTTAEQLGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKL 60
QY 61 HTLASGLSPGLRFGCTSTDFLIFNPNKDSWEEKVLSFEQAKDVCSEWSPAVVPKLL 120
DB 61 HTLASGLSPGLRFGCTSTDFLIFNPNKDSWEEKVLSFEQAKDVCSEWSPAVVPKLL 120
QY 121 TOWPLOEKLLLAHSHWKKHNTTITRSTLDILHTFASSSGFRVFGNALLRRAGLOWDS 180
DB 121 TOWPLOEKLLLAHSHWKKHNTTITRSTLDILHTFASSSGFRVFGNALLRRAGLOWDS 180
QY 181 SNAKQLLGCAQRSYNI SWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLYRH 240
DB 181 SNAKQLLGCAQRSYNI SWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLYRH 240
QY 241 AELYLGLDVQPKRHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPA 300
DB 241 AELYLGLDVQPKRHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPA 300
QY 301 TAIHDLVGLIVEATVPCKKWLGETSGAYGGAPOLSNYYVAGFMWLDKGLAARRGIDVV 360
DB 301 TAIHDLVGLIVEATVPCKKWLGETSGAYGGAPOLSNYYVAGFMWLDKGLAARRGIDVV 360
QY 361 MRQVSEFAGSYHLVDAGFKPLPDYWLISLLYKRLVGTQVLSQASVEQADARRPRVYLHCTNP 420
DB 361 MRQVSEFAGSYHLVDAGFKPLPDYWLISLLYKRLVGTQVLSQASVEQADARRPRVYLHCTNP 420
QY 421 RHPKYREGDVTLPALNLSNVTQSLQPKQWLSKSDYQYLLPHGKDSILSREVQLNGRL 480
DB 421 RHPKYREGDVTLPALNLSNVTQSLQPKQWLSKSDYQYLLPHGKDSILSREVQLNGRL 480
QY 481 QMVDDETLPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523
DB 481 QMVDDETLPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523

RESULT 2 SPAJD
Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6;
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0 (0).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AE8B13F07451684 CRC64;
```

SQ

```
Query Match 60.6%; Score 1653; DB 2; Length 574;
Best Local Similarity 60.7%; Pred. No. 8.4e-12;
Matches 315; Conservative 86; Mismatches 116; Indels 2; Gaps 2;
QY 6 LLVLLVLLVPPRTTAEQLGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKLHTLAS 65
DB 57 LVQCILAAQAEVDVELEFSTQRPDLHLVSPFSITIDANLATDPRFLTFGLSPKULRALAR 116
QY 66 GLSPGFLRFGCTSTDFLIFNPNKDSWEEKVLSFEQAK-DVCEAWPSPAVVPKLLLTQWP 124
DB 117 GLSPAYLRFGGTKTDFLIFDPKKEPSHEERSYKWSQVNHDCRSGAIPAVVVRRLQVEMP 176
QY 125 LQEKLLLAHSHWKKHNTTITRSTLDILHTFASSSGFRVFGNALLRRAGLOWDSNAK 184
DB 177 FQEQLLLRQEQVQKEFKNSYSSRSDMLYTPARCGLDLIFGLNALLRTADFRWNSSNAQ 236
QY 185 QLLCYCAQRSYNI SWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLYRHAEL 244
DB 237 LLNLYCSSKNYDLSWELGNEPNSFWKKAHISIDGLQLGEDYIELRKLKKSTL-KNVKLY 295
QY 245 GLDVGQPKRHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPATAIH 304
DB 296 GPDVGQPRGKTVKLLRSFLKAGGEVIDSVTWHYYLNGRIATKEDFLSPDVLDTFILSVQ 355
QY 305 DVLGIVEATVPCKKWLGETSGAYGGAPOLSNYYVAGFMWLDKGLAARRGIDVVMQV 364
DB 356 KILQVVEETRPCKKWLGETSSAYGGAPLLSNFAAGFMWLDKGLSAQMGIEVVMQV 415
QY 365 SFGAGSYHLVDAGFKPLPDYWLISLLYKRLVGTQVLSQASVEQADARRPRVYLHCTNPRPK 424
DB 416 FFGAGNYHLVDKNEPEPLPDYWLISLLFKKLVGSKVLMARVGPDRSKLRLVHCTNINHR 475
QY 425 YREGDVTLPALNLSNVTQSLQPKQWLSKSDYQYLLPHGKDSILSREVQLNGRLQWVD 484
DB 476 YQEGDLTYALNLYNVTKHKLPLYQLFNKVPDKYLVIPLPGGGLSKSVQLNGQALKWVD 535
QY 485 DETLPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523
DB 536 DQTLPALTEKPRPGSSGLPAPFSYGFYVIRNAKVAACL 574

RESULT 3
Q333X7 9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCB8BD99DE CRC64;

Query Match 60.4%; Score 1649; DB 2; Length 574;
Best Local Similarity 60.7%; Pred. No. 1.7e-12;
Matches 315; Conservative 85; Mismatches 117; Indels 2; Gaps 2;
```

```
QY 6 LVLVLLAVPPRTAELQGLREPICAVSPFLSLTLDASLARDPRFVALLRHPKHLTLAS 65
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 LVQCILAAQAEVDVELEFSTQRPHLVSPSPSLTIDANLTPRFLTFLGSPKRLARAL 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 66 GLSPGFLRFGGTSTDFLIFPNKDKSTWEEKVLSFQAK-DVCEAWPSPFVPKLLLTQMP 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 GLSPAYLRFGGTKTDFLIFPKPEPSHEERSYMKSQVNHDICRSGAIPVAVVRLEQWEP 176
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 LOEKLLLAHSWKHKVTTTTRSTLDLHFPASSGFLRVGLNALLRAGLQWDSNAK 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 FQEQLLREQYQKEFKNSTYSRSSVDMLYTFARGSGLDLIFGLNALLRTADFRWNSSNAQ 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 QLLGYCAQRSYNTLSWELGNPNPSFRKSGICIDGFGQLGRDPVHLRQLLSQHPLYRHAELY 244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 LLLNYGCKNYDLSWELGNPNPSFWKKAHISIDQLGQEDYIEURKLKRLSTL-KNKVLY 295
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 GLDVGPGRKHTQHLRLSRFMKSGGKAIDSVTHHHYVNGRSATREDFLSPVLDTSFATAIH 304
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 GPDVGQGRGTVKLLRSFLKAGEVIDSVTHHHYVNGRATKEDFLSPVLDTFILSVQ 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 DVLGIVEATVPGKKVWLGTSAGYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 KILQVVEETRPGKKVWLGTSAGYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 415
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 SFAGSGVHLVDAGFKPLPDYWLSSLLYKRLVGRTRVLQASVQDARRPRVYLHCTNPRHPK 424
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 FFGAGNYHLVDKNPEPLPDYWLSSLLYKRLVGRTRVLQASVQDARRPRVYLHCTNPRHPK 475
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 YRGEDVTLFALNLSNTOSLQPKLWSKSDVQVLLPHGKDSLTLSEVOLNGLLQWMD 484
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 476 YQEGDUTLALNLYNNVTHKULPQLEKNPDVKYLVKPLGPGGLSKSVQLNGOALKWMD 535
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 DETLPALHMAAPGSTGLPAPFSYGFYVIRNAKAIACI 523
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 DQTLPALTEKPLGSGSLPAPFSYGFYVIRNAKVAACL 574
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 4
HPSE HUMAN STANDARD; PRT; 543 AA.
AC Q9Y2E1; Q53GE5; Q9UL39;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1999, sequence version 1.
DE 07-FEB-2006, entry version 27.
DE Heparanase precursor [EC 3.2.-.-] (Heparanase-1) (Hpa1) (Endo-
DE glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
DE subunit).
GN Name=HPSE; Synonym=HEP, HPA, HPA1, HPR1, HPSE1, HSE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
RA Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and functional expression of a human heparanase gene.";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
[2]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
RP SEQUENCE OF 158-168; 326-337 AND 447-491.
RC TISSUE=Embryonic fibroblast;
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression.";
RL J. Biol. Chem. 274:24153-24160(1999).
[3]
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
```

```
RX PubMed=10395325; DOI=10.1038/10518;
RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,
RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Pecker I.;
RT "Mammalian heparanase: gene cloning, expression and function in tumor
RT progression and metastasis.";
RL Nat. Med. 5:793-802(1999).
[4]
RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parieh C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
[5]
RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RA "Heparanase expression in invasive trophoblasts and acute vascular
RT damage.";
RL Glycobiology 10:467-475(2000).
[6]
RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
RP PubMed=11547900; DOI=10.1023/A:1011375624902;
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
RT "Molecular properties and involvement of heparanase in cancer
RT progression and mammary gland morphogenesis.";
RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
[7]
RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
RC TISSUE=Placenta;
RX PubMed=12713442; DOI=10.1042/BJ20030318;
RA McKenzie E., Young K., Hancock M., Bennett J., Bhaman M., Felix R.,
RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
RA Parekh R., Stubberfield C.;
RT "Biochemical characterization of the active heterodimer form of human
RT heparanase (Hpa1) protein expressed in insect cells.";
RL Biochem. J. 373:423-435(2003).
[8]
RN NUCLEOTIDE SEQUENCE [MRNA].
RP Pinhal M.A., Semedo P.;
RA "Cloned heparanase from MCF-7 cells.";
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
[9]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Small intestine;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
[10]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RP TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Scheetz T.E.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Spalacidae; Spalacinae; Spalax.
 OX NCBI_TaxID=164323;
 [1]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 cloning and identification of a novel splice variant.";
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
 [2]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 cloning and identification of a novel splice variant.";
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AM085490; CAJ30017.1; -; mRNA.
 SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

 Query Match 60.2%; Score 1641; DB 2; Length 574;
 Best Local Similarity 60.5%; Pred. No. 7.5e-121;
 Matches 314; Conservative 84; Mismatches 119; Indels 2; Gaps 2;
 QY 6 LVLVLLAVPRRTAELQGLREPIGAVSPAFSLTLDASLARDPRFVALLRHPKLUHTLAS 65
 DB 57 LVQCILAAQAQADVVELEFSTQRLHLVSPSFLSIDANLATDPRFLTGLGSPKLALAR 116
 QY 66 GLSPGFLRFQGTSTDFLIENPNKDSWEEKVSEFOAK-DVCEAWPSFAVVPKLLLTQWP 124
 DB 117 GLSPAYLRFGGTTDFLFPKPEPSHEERSYWKSVQVNHIDCRSGAIPAVVVRRLQVWP 176
 QY 125 LQEKLLAAHSWKHKNTTITRSLDILHTFASSSGFLVFGNLALRRAGLQWDSNAK 184
 DB 177 FQEQLLREQYQKDFKNSTYSRSSVDMLYTFARCSGLDLIFGLNALLRTADFRWNSSNAQ 236
 QY 185 QLLGCAQRSYNTSWELGNPNFRKSGICIDGFLGRDFVHLRQLLSQHPLYRHAELY 244
 DB 237 LLLNYCCKNYDISWELGNPNFRKSGAHSIDGLQGEDYIELHKLRLKSTL-KNVKLY 295
 QY 245 GLDVGQPRKHTQHLRSFMKSGKAIDSVTHHYVNGRSATREDFLSPVLDSEFATAIH 304
 DB 296 GPDVGQPRGKTVKLLRSFLKAGGEVIDSVTHHYVNGRIATKEDFLSPDVLDTFLSVQ 355
 QY 305 DVLGIVEATVPKKVWLGETSGAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364
 DB 356 KILQVVEETRPKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLSAGMGLVVMRQV 415
 QY 365 SFGAGSYHLVDAGFKPLPDYWLSSLYKRLVGTTRVLQASVEQADARRPRVYLHCTNPRHPK 424
 DB 416 FFGAGNYHLVDKNFEPLPDYWLSSLFKL VGSKVLMARVGPDRSKLRVYLHCTNINHRP 475
 QY 425 YREGDVTLPALNLSNTQSLQPKLWSKVDQYLLPHGKDSILSREVQLNGELLOMVD 484
 DB 476 YQEGDLTLALNLYNVTKHLKLPYQLFNKPDYLVKPLPGGLLSKSVQLNGQALQWVD 535
 QY 485 DETLPALHEMALAPGSTLGLPAPFSYGFYVIRNAKAIACI 523
 DB 536 DQTLPALTEKPLRPGSSGLPAPFSYGFYVIRNAKVAACL 574

 RESULT 6
 Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
 AC Q333X8;
 DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
 DT 06-DEC-2005, sequence version 1.
 DT 07-FEB-2006, entry version 3.

DE Heparanase.
 GN Name=hpa.
 OS Spalax golani.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muroidae; Spalacidae; Spalacinae; Spalax.
 OX NCBI_TaxID=191382;
 [1]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 cloning and identification of a novel splice variant.";
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
 [2]
 RN NUCLEOTIDE SEQUENCE.
 RC TISSUE=Kidney;
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
 cloning and identification of a novel splice variant.";
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AM085491; CAJ30018.1; -; mRNA.
 SQ SEQUENCE 574 AA; 64555 MW; 48BEPEC7D0BCB34 CRC64;

 Query Match 60.1%; Score 1640; DB 2; Length 574;
 Best Local Similarity 60.3%; Pred. No. 8.9e-121;
 Matches 313; Conservative 86; Mismatches 118; Indels 2; Gaps 2;
 QY 6 LVLVLLAVPRRTAELQGLREPIGAVSPAFSLTLDASLARDPRFVALLRHPKLUHTLAS 65
 DB 57 LVQCILAAQAQADVVELEFSTQRLHLVSPSFLSIDANLATDPRFLTGLGSPKLALAR 116
 QY 66 GLSPGFLRFQGTSTDFLIENPNKDSWEEKVSEFOAK-DVCEAWPSFAVVPKLLLTQWP 124
 DB 117 GLSPAYLRFGGTTDFLFPKPEPSHEERSYWKSVQVNHIDCRSGAIPAVVVRRLQVWP 176
 QY 125 LQEKLLAAHSWKHKNTTITRSLDILHTFASSSGFLVFGNLALRRAGLQWDSNAK 184
 DB 177 FQEQLLREQYQKDFKNSTYSRSSVDMLYTFARCSGLDLIFGLNALLRTADFRWNSSNAQ 236
 QY 185 QLLGCAQRSYNTSWELGNPNFRKSGICIDGFLGRDFVHLRQLLSQHPLYRHAELY 244
 DB 237 LLLNYCCKNYDISWELGNPNFRKSGAHSIDGLQGEDYIELHKLRLKSTL-KNVKLY 295
 QY 245 GLDVGQPRKHTQHLRSFMKSGKAIDSVTHHYVNGRSATREDFLSPVLDSEFATAIH 304
 DB 296 GPDVGQPRGKTVKLLRSFLKAGGEVIDSVTHHYVNGRIATKEDFLSPDVLDTFLSVQ 355
 QY 305 DVLGIVEATVPKKVWLGETSGAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364
 DB 356 KILQVVEETRPKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLSAGMGLVVMRQV 415
 QY 365 SFGAGSYHLVDAGFKPLPDYWLSSLYKRLVGTTRVLQASVEQADARRPRVYLHCTNPRHPK 424
 DB 416 FFGAGNYHLVDKNFEPLPDYWLSSLFKL VGSKVLMARVGPDRSKLRVYLHCTNINHRP 475
 QY 425 YREGDVTLPALNLSNTQSLQPKLWSKVDQYLLPHGKDSILSREVQLNGELLOMVD 484
 DB 476 YQEGDLTLALNLYNVTKHLKLPYQLFNKPDYLVKPLPGGLLSKSVQLNGQALQWVD 535
 QY 485 DETLPALHEMALAPGSTLGLPAPFSYGFYVIRNAKAIACI 523
 DB 536 DQTLPALTEKPLRPGSSGLPAPFSYGFYVIRNAKVAACL 574

 RESULT 7
 HPSE_BOVIN
 ID HPSE_BOVIN
 AC Q9MY0;
 STANDARD; PRT; 545 AA.

DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
GN Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=2117669; PubMed=1127787;
RT Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RA "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AF281160; AAF87301.2; -; mRNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_Hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL. 1 37 By similarity.
FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
FT /FTID=PRO_0000042256.
FT PROPEP 112 159 Linker peptide.
FT /FTID=PRO_0000042257.
FT CHAIN 160 545 Heparanase 50 kDa subunit (By
FT similarity).
FT /FTID=PRO_0000042258.
FT REGION 160 164 Heparin/HS-binding (Potential).
FT REGION 272 282 Heparin/HS-binding (Potential).
FT ACT_SITE 227 227 Proton donor (Potential).
FT ACT_SITE 345 345 Nucleophile (Potential).
FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD85B933 CRC64;

Query Match 59.5%; Score 1624; DB 1; Length 545;
Best Local Similarity 58.1%; Pred. No. 1.5e-119;
Matches 311; Conservative 93; Mismatches 115; Indels 16; Gaps 4;
Qy 3 VLLLLVLL-----LAVPPRTAEQLGLREPIGAVSPAFSLTLDASLARDPR 50
:||||| :
Db 13 LLLLLPLGLGPCSPGTPAAAAPADAAAELEFFTERPLHLVSPAFSLTIDANLATDPR 72
:||||| :
Qy 51 FVALLRHPKLHTLASGLSPGFLRFGTSTDFLIINPNKDSWEEKV--LSFQAKDVCEA 108
:||||| :
Db 73 FFFFLGSSKLRLTARGLAPAYLRFEGNGKGDFLIPDKKEPAFEERSYWLQ--SNQDICKS 131
:||||| :
Qy 109 WPSFAVVPKLLLTOWPLOEKLLLAHESWKHKHNTITRSTLDILHTPASSGSFRLVFGLN 168
:||||| :
Db 132 GSIPSDVEEKLRLWEPFQEQVLLREYQKQKFTNSTYSRSSVDMLTYTFASCGLNLIFGVN 191
:||||| :
Qy 169 ALLRRAGLQDSSNAKOLLGYCAORSYNIWSLGNENPSFRKSGICICIDGFGOLGRDFVHL 228
:||||| :
Db 192 ALLRTTMDHWDSSNAQLLDYCSSKNYNIWSLGNENPSFORKAGIFINGQLGEDFIEF 251
:||||| :
Qy 229 ROLLSQHLRYHAELYGLDVGQPKRHTQHLRLSPFMKSGKKAIDSVTWHYYVNGRSATRE 288
:||||| :
Db 252 RKLGLK-SAFKNALYGPDIGQPRNTVYMLKSLKAGGEVIDSVTWHYYVNGRIATKE 310
:||||| :
Qy 289 DFLSPVELDSFATAIHVDVLGIVEATVPGKKVWLGTSAYCGGAPOLSNVTVAGFMWLDK 348
:||||| :
Db 311 DFLNPDLDTFISVQVKTIRIVEKIRPLKVKWLGTSAYCGGAPFLSNVTVAGFMWLDK 370
:||||| :
Qy 349 LGLAARGIDVWROVSFGAGSHLVDAQFKPLDYLKRLVGRVLOASVEQADA 408
:||||| :
Db 371 LGLSARMGIEVWVKQVLFAGNYHLVDGNFELPDYWLSLFKLVGNKVLMAVKGDPDR 430
:||||| :
Qy 409 RRPVYLHCTNPRHPKYREGDVTFLALNLSNVTSLOLPKOLWSKSDVDYLLPHGKDSI 468
:||||| :
Db 431 SKFRVYLHCTNKHPRYKEGDLTIYALNHNVTKHLELPHLEKNQVDKYLIPKSGTDGL 490
:||||| :
Qy 469 LSRVQLNGRLQLQWDDDETLPALHEMALPGSTGLPAFSGYGFVIRNAKAIACI 523
:||||| :
Db 491 LSKSVQLNGQLKMWDEQTLPALTEKPLHPGSSGLMPPFPFSGFFVIRNAKAAACI 545
:||||| :
RESULT 8
ID HPSE MOUSE STANDARD; PRT; 535 AA.
AC Q6YGL; Q8K3K3;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=HPse; Synonyms=Hpa;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=JUL/J; TISSUE=Spleen;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RA "Cloning of mammalian heparanase, an important enzyme in tumor
RA invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
RL [2]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
RP GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND
RP SUBUNITS.
RC STRAIN=FVB; TISSUE=Embryo;
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Mao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;

RT "Cloning, expression, and purification of mouse heparanase."; [3]
 RN PROTEIN SEQUENCE [LARGE SCALE MRNA].
 RP NUCLEOTIDE SEQUENCE [MRNA] AND ENZYME REGULATION.
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
 RA Gong P., Jemth P., Galvis M.L.E., Vlodaysky I., Horner A., Lindahl U.,
 Li J.-P.;
 RA "Processing of macromolecular heparin by heparanase."; [4]
 RL J. Biol. Chem. 278:35152-35158 (2003).
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
 RX PubMed=16141072; DOI=10.1126/science.1112014;
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
 Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
 Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
 Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
 Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
 Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
 Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
 Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
 di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
 Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
 Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
 Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
 Hill D., Huminecki L., Iacono M., Ikeo K., Iwano A., Ishikawa T.,
 Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
 Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
 Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
 Litani S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
 Matuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
 Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
 Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
 Okazaki Y., Ohtsundo V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
 Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
 Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
 Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
 Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
 Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
 Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
 Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yang K.,
 Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
 Grimmond S.M., Teasdale R.D., Liu E.T., Bruscia V., Quackenbush J.,
 Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
 Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
 Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
 Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
 Nishio T., Okada M., Plesey C., Shibata K., Shiraki T., Suzuki S.,
 Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,
 Hayaishizaki Y.;
 RL "The transcriptional landscape of the mammalian genome."; Science 309:1559-1563 (2005).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans. Also implicated in the extravasation of leukocytes and tumor cell lines. Contributes to metastasis and angiogenesis (by similarity).
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and magnesium (by similarity). Inhibited by laminarin sulfate and, to a lower extent, by heparin and sulfamin.
 CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 5;
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa subunits, the proteolytic products.
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted. Secreted, internalised and transferred to late endosomes/lysosomes as a proheparanase. In lysosomes, it is processed into the active form, the heparanase. The uptake or internalisation of proheparanase is mediated by HSPGs. Heparin appears to be a competitor and retain proheparanase in the extracellular medium (by similarity).

CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form leads to the generation of a linker peptide, 8 kDa and 50 kDa product. The active form, the 8/50 kDa heterodimer, is resistant to degradation. Complete removal of the linker peptide appears to be a prerequisite to the complete activation of the enzyme (by similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears to be essential for its solubility.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC EMBL: AF359507; AAQ15188.1; -; mRNA.
 CC EMBL: AY077467; AAL76083.1; -; mRNA.
 CC EMBL: AY151051; AAN41636.1; -; mRNA.
 CC EMBL: AK040471; BAC30600.1; -; mRNA.
 CC EMBL: AK154628; BAE32725.1; -; mRNA.
 CC Ensembl: ENSMUSG0000035273; Mus musculus.
 CC MGI: MGI:1343124; Hpse.
 CC GO: GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
 CC InterPro: IPR005199; Glyco_hydro_79_N.
 CC Pfam: PF03662; Glyco_hydro_79n; 1.
 CC KEGG: Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
 CC Magnesium; Membrane; Signal.
 CC SIGNAL 1 27 By similarity.
 CC CHAIN 28 101 Heparanase 8 kDa subunit.
 CC PROPEP 102 149 /FTID=PRO 0000042263.
 CC CHAIN 150 535 Linker peptide (By similarity).
 CC /FTID=PRO 0000042264.
 CC Heparanase 50 kDa subunit.
 CC /FTID=PRO 0000042265.
 CC REGION 150 154 Heparin/HS-binding (By similarity).
 CC REGION 262 272 Heparin/HS-binding (By similarity).
 CC ACT SITE 217 217 Proton donor (Potential).
 CC ACT SITE 335 335 Nucleophile (Potential).
 CC CARBOHYD 154 154 N-linked (GlcNAc...).
 CC CARBOHYD 192 192 N-linked (GlcNAc...).
 CC CARBOHYD 209 209 N-linked (GlcNAc...).
 CC CARBOHYD 230 230 N-linked (GlcNAc...).
 CC CARBOHYD 451 451 N-linked (GlcNAc...).
 CC CONFLICT 206 206 K -> R (in Ref. 3).
 CC CONFLICT 212 212 W -> S (in Ref. 3).
 CC CONFLICT 230 232 NGS -> DGL (in Ref. 1, 2 and 4).
 CC CONFLICT 335 335 E -> K (in Ref. 3).
 CC CONFLICT 342 342 G -> A (in Ref. 3).
 CC CONFLICT 455 455 Y -> H (in Ref. 1, 2 and 4).
 CC CONFLICT 531 531 V -> I (in Ref. 1, 2 and 4).
 CC SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;
 CC
 CC Query Match 58.7%; Score 1602; DB 1; Length 535;
 CC Best Local Similarity 58.8%; Pred. No. 8.3e-118;
 CC Matches 315; Conservative 82; Mismatches 125; Indels 14; Gaps 4;
 CC
 CC QY 1 MLVLLLVL-----LLAVPPRTA-----ELQGLREPGVAPAFSLTLTLDASLARD 48
 CC DQ 1 MLRLLLLMGPLGALGAPAGTAPTDVVDLEFYTKRPLRSVSPFLSITDASLATD 60
 CC
 CC QY 49 PRFVALLRHPKLTHTLGLSPGLRFGCTSTDFLIFNPNKSDTWEEKVLSEFOAK-DVCE 107
 CC DB 61 PRFTFLGSPRLRALRGLSPAYLRFPGTKTDFLIFDPDKPTSEBSRWKQVNHDIR 120
 CC
 CC QY 108 AWPSPVAVPKLLLTQWPLQEKLLLAHKKHNTTITRSTLDILHTFASSGFRLVFGL 167
 CC DB 121 SEPSAAVLRKLOVENPQELLRLLEQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGL 180
 CC
 CC QY 168 NALLRAGLQWDSNNAKOLLGYCAQRSYNIWELGNENPNSFFKSGICITDGLQGRDFVH 227
 CC DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNIWELGNENPNSFWKHAHLINGSQLGDFVE 240
 CC
 CC QY 228 LRQLLSQHPLYRHAELLYGLDVQPRKHTQHLLRSFMSKGGKALDSVTHHHYVNGSATR 287
 CC DB 241 LHKLL-QRSFAFNAKLGYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATK 299

QY 288 EDFLSEVLDSFATAIHVDVIGVEATVPGKKVWLGTSAYGGAPOLSNYYVAGFMWLD 347
 Db 300 EDFLSSDLVTLFSLVQKILKVTKEITPGKKVWLGTSAYGGAPLSTNTFAAGFMWLD 359
 QY 348 KGLGAARRGIDVVMQVSGAGSVHLVDAGFKPLPDYWLKLLKRLVGTIVLQASVEQAD 407
 Db 360 KGLSQAOMGIEVVMQVQFFGAGNTHLVDPENPELPDYWLKLLKRLVGTIVLQASVEQAD 419
 QY 408 ARPRVYLHCTNPRHPKRYEGDVTFLFALNLSNVTQSLQPLKQMSKSDVDQVLLPHGKDS 467
 Db 420 RSKLRVYLHCTNVYHPRYQSGDLTLVNLHNTVKYLKVPPLPKRPVDPYLLKPSGPDG 479
 QY 468 ILRSVQLNGRLQWVDDELTPALHEWALAPGSTGLPAPFSYGFYVIRNAKAIACI 523
 Db 480 LLKSKVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLPFSYGFYVIRNAKAIACI 535

RESULT 9
 HPSE_RAT
 ID HPSE RAT STANDARD; PRT; 536 AA.
 AC Q71R1; Q9QZF8;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 05-JUL-2004, sequence version 1.
 DT 07-MAR-2006, entry version 11.
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
 GN Name=Hps; Synonyms=Hep;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RX [1]
 RP NUCLEOTIDE SEQUENCE [MRNA].
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
 RA Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis."
 RL Nat. Med. 5:803-809 (1999).
 [2]
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
 RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;
 RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
 RT "Characterization of heparanase from a rat parathyroid cell line."
 RL J. Biol. Chem. 277:32459-32465 (2002).
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Contributes to metastasis and angiogenesis
 CC (By similarity).
 CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
 CC extent, by heparin and sulfamin (By similarity). Activated by
 CC calcium and magnesium. Inhibited by EDTA.
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products (By similarity).
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium
 CC (By similarity).
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme (By
 CC similarity).
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears

CC to be essential for its solubility (By similarity).
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
 CC
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC
 CC -----
 CC EMBL; AF359508; AAQ15189.1; -; mRNA.
 DR EMBL; AF184967; AAF04563.1; -; mRNA.
 DR RGD; 61969; Hps.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
 KW Signal.
 FT SIGNAL 1 28 By similarity.
 FT CHAIN 29 102 Heparanase 8 kDa subunit.
 FT /FTID=PRO_0000042266.
 FT PROPEP 103 150 Linker peptide (By similarity).
 FT /FTID=PRO_0000042267.
 FT CHAIN 151 536 Heparanase 50 kDa subunit.
 FT /FTID=PRO_0000042268.
 FT REGION 151 155 Heparin/HS-binding (By similarity).
 FT REGION 263 273 Heparin/HS-binding (By similarity).
 FT ACT_SITE 218 218 Proton donor (Potential).
 FT ACT_SITE 336 336 Nucleophile (Potential).
 FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).
 FT CONFLICT 15 15 G -> R (in Ref. 2).
 FT CONFLICT 227 227 H -> Q (in Ref. 2).
 FT CONFLICT 350 350 D -> N (in Ref. 2).
 SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;
 Query Match 57.6%; Score 1572; DB 1; Length 536;
 Best Local Similarity 58.9%; Pred. No. 2e-115;
 Matches 302; Conservative 87; Mismatches 118; Indels 6; Gaps 3;
 QY 14 PPRRTAEQLQLREPIGAVSPAFSLTLDSALDRPRFVALLRHPKHLTASGLSPGFLR 73
 Db 27 PTKDQVDELYFTKRLFSQVSPSLITIDASLATDPRFTLFGSPRLARALGSLPAYLR 86
 QY 74 FGGTSTDFLIPNKNOSTWEEKVSEFOAK---DVCAMPSFAVVPKLLLTQWLPQEKLL 130
 Db 87 FGGTKTDFLIFDPNKEPTSEER--SYWQSDNNDICGSERVSADVLKRLQMEWPFQELL 144
 QY 131 LAEHSKKKHNVTITRSTDLIHTFASSGFRVFGNLALRRAGLQWDSNAKQLLYGC 190
 Db 145 LRSQYQREKKNSTYSSSDMLYSFAKSRDLDLIFGLNALLRTPDLRWNSNAQLLLNYC 204
 QY 191 AQRSYNISWELGNEPNSFRKSGICIDGOLFGRDFVHLRQLLSQHPLYRHAELYGLDVQ 250
 Db 205 SSKGYNISWELGNEPNSFWKKAHISIDGLQGEDFVELHKL--QKSAFQNAKLYGPDIG 263
 QY 251 PRKHTQHLRSFMKSGKKAIDSVTHHHYVNGRSATREDFLSPVELDSFATAIHDLVGI 310
 Db 264 PRGKTVLLRSFLKAGGEVIDSLTWHYILNGRVATKEDFLSDVLDTFILLSVQKILVT 323
 QY 311 EATVPGKKVWLGTSAYGGAPOLSNYYVAGFMWLDKGLAARRGIDVVMQVSGAGS 370
 Db 324 KEMTPGKKVWLGTSAYGGAPLSDTFAAGFMWLDKGLSAQIGIEVVMQVFFGAGN 383
 QY 371 YHLVDAGFKPLPDYWLKLLKRLVGTIVLQASVEQADARRPRVYLHCTNPRHPKRYEGD 430
 Db 384 YHLVDENPELPDYWLKLLKRLVGTIVLQASVEQADARRPRVYLHCTNPRHPKRYEGD 443
 QY 431 TLFALNLSNVTQSLQPLKQMSKSDVDQVLLPHGKDSILSREVQLNGRLQWVDDELTPA 490
 Db 444 TLVYLNLHNTVKYLKVPPLPKRPVDPYLLKPSGPDG 503
 QY 491 LHEWALAPGSTGLPAPFSYGFYVIRNAKAIACI 523
 Db 504 LTEKPLPAGSSLSVPAPFSYGFYVIRNAKAIACI 536

[illegible]

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.:
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;
 RG NIH MGC Project;
 RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; BC112356; AA112357.1; -; mRNA.
 SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;

Qy	1	MLVLLLVLLLA-----VPPRTABLQ-----LGLREPIGAVSPAFSLITLDAS	44
Db	27	LYLALLHLSSQAGDRPLPDVDAAGLKEKTLILLDVSTKNPRTVNFNLSLQLDPS	86
Qy	45	LARDPRFVALLRHPKLHTLGLSPGFIRPGTSTDFLIIP-----NPNKD-----S	90
Db	87	IIHD-GWLDLFSSKRLVLTARGLSPAFIRFGGKRTDFLQFNLRNPAKSRGGPGPDYYLK	145
Qy	91	TWEKVLSEFOAKDV--CEAWPSFVVPKLLLTQWPLOEK-----LLLAHSWKHGK	140
Db	146	NYEODIVRSQVVALDKQCK-----IAQHPDVM-----ELQREKAAQMHVLVLKEQFSNTYS	198
Qy	141	NTTITRSTDLHTFPASSGFRFLVFLGNALLRRAGLOWDSNAKQLLGYCAQRSYNISWE	200
Db	199	NLIILARSCLKYNFADCSGLHLIFALNALRPNNSWNSSALSLLKYSASKYNISWE	258
Qy	201	LGNPNSFRKSGICIDQFGLGRDFVHURLQLLSOHLRYHAEYLGLDVGQPRKHTQHLLR	260
Db	259	LGNEPNRYTRHGRAVNGSLQGDYIQLKSLIQLPIRYSRASLYGPNIGRPRKNVIALLD	318
Qy	261	SPMKSGKAIQSVTWHYVNGRSATREDFLSPVILDSFAITHDVLGIVEATVPGKKVW	320
Db	319	GFMKVAGSTDAVTAWQHICYIDGRVVKVMDFLKTRLLDLSOIRKIQKVNTYTPGKKIW	378
Qy	321	LGETGSYGGGAPOLSNITYAVGFWMWLDKGLAARRGIDVNRQVSFGAGSVHLVDAGFKP	380
Db	379	LEGVVYTSAGGTNNLSDSYAGFLWNTLGLMANQGDIVLRHSHFFDGYHNLVDQFNFP	438
Qy	381	LPDYWLSLLYKRLVGTGRVLQASVQADAR-RP-----RVYLHCTNPRHPKRYREGDVT	431
Db	439	LPDYWLSLLYKRLIGPKVLAVHVAGLQRPGRVIRDKLRIYAHCTNHHNHNVRGSIIT	498
Qy	432	LPALNLSNVTOSLOLPKOLWSKSDVOYLLPHGKDSILSRVOLNRLLOWVDDETLPAL	491
Db	499	LFIIHLHRSRKIKLXTGLTRDKLVHQYLLQYGGEGLSKSVQLNGQPLVMVDDGTPEL	558
Qy	492	HEMALAPGSTLGLPAFSGYGFVIRNAKAIC	522
Db	559	KPREFRAGRTLVIPPTVNGFFVKKVNALAC	589

Query Match 37.7%; Score 1028.5; DB 2; Length 592;
 Best Local Similarity 40.6%; Pred. No. 2.4e-72;
 Matches 232; Conservative 83; Mismatches 199; Indels 57; Gaps 11;

RESULT 14
 Q4TB80_TETNG
 1D Q4TB80_TETNG PRELIMINARY; PRT; 597 AA.
 AC Q4TB80;
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
 DT 19-JUL-2005, sequence version 1.
 DT 07-FEB-2006, entry version 4.
 DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
 GN ORFNames=GSTNG000386001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

[illegible]

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 69.589 Seconds

(without alignments)
65.702 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CTWTDNPRYK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

- A_Geneseq_8:*
1: Geneseqpl980s:*
2: Geneseqpl990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*
10: Geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	10	8	ADR88212 Human hep
2	61	100.0	10	8	ADT78179 Functiona
3	61	100.0	10	8	AEA42428 Human hep
4	61	100.0	15	9	ADU71246 Human hep
5	61	100.0	15	9	ADU70983 Human hep
6	61	100.0	15	9	ADU71087 Human hep
7	61	100.0	15	9	ADU71245 Human hep
8	61	100.0	386	8	ADR88207 Human mat
9	61	100.0	386	8	ADT78174 45kDa sub
10	61	100.0	386	9	ADY27057 Heparanas
11	61	100.0	386	9	ADZ18995 Human hep
12	61	100.0	386	9	AEA42423 Human mat
13	61	100.0	460	9	ADY27061 Heparanas
14	61	100.0	486	9	AE887589 Human hep
15	61	100.0	492	9	ADZ18996 Hep106 co
16	61	100.0	493	9	AE887562 Human hep
17	61	100.0	495	9	ADZ18999 Hep109 co
18	61	100.0	497	9	AE887587 Human hep
19	61	100.0	501	9	ADZ19000 HepGS3 co
20	61	100.0	507	9	ADZ19005 HepGS6 co
21	61	100.0	508	9	ADY27058 Human ina
22	61	100.0	526	9	ADZ19006 Hepfyalur
23	61	100.0	527	5	ABB07815 Chicken s

24	61	100.0	527	7	ABW02018	Abw02018 Chimeric
25	61	100.0	527	8	ADO63825	ADO63825 Chimeric
26	61	100.0	527	8	ADO63827	ADO63827 Chimeric
27	61	100.0	527	8	ADO63826	ADO63826 Chimeric
28	61	100.0	527	9	ADZ19004	Adz19004 HepGS4 co
29	61	100.0	530	2	AAZ34173	AAZ34173 Human pro
30	61	100.0	532	2	AAZ17083	AAZ17083 Seq ID No
31	61	100.0	543	2	AAZ02345	AAZ02345 A human h
32	61	100.0	543	2	AAZ17082	AAZ17082 Human hep
33	61	100.0	543	3	AAZ57590	AAZ57590 Human hep
34	61	100.0	543	3	AAZ08849	AAZ08849 Amino aci
35	61	100.0	543	3	AAZ52990	AAZ52990 Human aci
36	61	100.0	543	4	AAZ97635	AAZ97635 Human hep
37	61	100.0	543	4	AAZ86206	AAZ86206 Human hep
38	61	100.0	543	4	AAZ88361	AAZ88361 Human mem
39	61	100.0	543	5	ABB07813	ABB07813 Human hep
40	61	100.0	543	7	ADD18950	ADD18950 Human dis
41	61	100.0	543	7	ADG88800	ADG88800 Human hpa
42	61	100.0	543	8	ADL16379	ADL16379 Human hep
43	61	100.0	543	8	ADK52086	ADK52086 Human ato
44	61	100.0	543	8	ADM48716	ADM48716 Human hpa
45	61	100.0	543	8	ADM48759	ADM48759 Human hpa
46	61	100.0	543	8	ADN05074	ADN05074 Antipeori
47	61	100.0	543	8	ADN04902	ADN04902 Antipeori
48	61	100.0	543	8	ADO63831	ADO63831 Human hep
49	61	100.0	543	8	ADO63824	ADO63824 Human hep
50	61	100.0	543	8	ADO63823	ADO63823 Human hep
51	61	100.0	543	8	ADO63832	ADO63832 Human hep
52	61	100.0	543	8	ADO63822	ADO63822 Human hep
53	61	100.0	543	8	ADQ80372	ADQ80372 Heparanas
54	61	100.0	543	8	ADR88210	ADR88210 Human pre
55	61	100.0	543	8	ADP25079	ADP25079 PRO polyp
56	61	100.0	543	8	ADT78177	ADT78177 Human hep
57	61	100.0	543	9	ADY27036	ADY27036 Human hep
58	61	100.0	543	9	ADY63087	ADY63087 Human clo
59	61	100.0	543	9	AEA42466	AEA42466 Human hep
60	61	100.0	543	9	AEA42426	AEA42426 Human hep
61	61	100.0	543	10	AE896848	AE896848 Human hep
62	61	100.0	545	6	ABP56822	ABP56822 Human hep
63	61	100.0	545	7	ADE16012	ADE16012 G-coupled
64	61	100.0	545	8	ADL93951	ADL93951 Human G-c
65	61	100.0	556	9	ADZ19010	ADZ19010 Heparanas
66	61	100.0	570	9	ADZ19008	ADZ19008 Heparanas
67	61	100.0	588	2	AAZ30124	AAZ30124 A human p
68	61	100.0	592	2	AAZ02346	AAZ02346 A human h
69	61	100.0	592	3	AAZ08850	AAZ08850 Amino aci
70	61	100.0	592	7	ADG88804	ADG88804 Human SK-
71	61	100.0	592	8	ADL16383	ADL16383 Human hep
72	61	100.0	592	8	ADM48720	ADM48720 Human SK-
73	61	100.0	592	9	AEA42461	AEA42461 Human hep
74	56	91.8	9	9	ADU70671	ADU70671 Human hep
75	56	91.8	15	9	ADU70848	ADU70848 Human hep
76	49	80.3	15	9	ADU70982	ADU70982 Human hep
77	49	80.3	545	9	ADY27034	ADY27034 Bovine he
78	47	77.0	9	9	ADU70619	ADU70619 Human hep
79	47	77.0	15	9	ADU71247	ADU71247 Human hep
80	44	72.1	9	9	ADU70744	ADU70744 Human hep
81	44	72.1	15	9	ADU71086	ADU71086 Human hep
82	44	72.1	819	8	ADN24558	ADN24558 Bacterial
83	44	72.1	858	8	ADN21798	ADN21798 Bacterial
84	43	70.5	487	8	ADN25773	ADN25773 Bacterial
85	43	70.5	639	3	AAZ74438	AAZ74438 Neisseria
86	43	70.5	639	3	AAZ74438	AAZ74438 Neisseria
87	43	70.5	639	3	AAZ74440	AAZ74440 Neisseria
88	43	70.5	648	6	ABP80492	ABP80492 N. gonorr
89	43	70.5	820	8	ADS27119	ADS27119 Bacterial
90	43	70.5	829	8	ADN26735	ADN26735 Bacterial
91	43	70.5	839	8	ADN26572	ADN26572 Bacterial
92	43	70.5	839	8	ADN26757	ADN26757 Bacterial
93	43	70.5	858	8	ADS24760	ADS24760 Bacterial
94	43	70.5	860	9	ADU81528	ADU81528 Aconitase
95	43	70.5	860	9	ADU81526	ADU81526 Aconitase
96	43	70.5	864	8	ADS26367	ADS26367 Bacterial

97	43	70.5	865	9	ADU81525	Adu81525	Aconitase	170	38	62.3	871	8	ADS24438	Bacterial
98	43	70.5	866	9	ADU81527	Adu81527	Aconitase	171	38	62.3	882	8	ADN25341	Bacterial
99	43	70.5	867	9	ADU81524	Adu81524	Aconitase	172	38	62.3	884	8	ADS21824	Bacterial
100	43	70.5	868	3	AAV93301	Aay93301	Amino aci	173	38	62.3	887	8	ADN24259	Bacterial
101	43	70.5	868	3	ADU81522	Adu81522	Aconitase	174	38	62.3	887	8	ADN25875	Bacterial
102	43	70.5	868	9	ADU81523	Adu81523	Aconitase	175	38	62.3	887	8	ADN28752	Bacterial
103	43	70.5	869	3	AAV93285	Aay93285	Amino aci	176	38	62.3	887	8	ADY24031	Plant ful
104	43	70.5	871	7	ABO80676	Abo80676	Pseudomon	177	38	62.3	889	5	ABB57134	Mouse isc
105	43	70.5	875	6	ADA333597	Ada333597	Acinetoba	178	38	62.3	889	8	ADN04279	Antipsori
106	41	67.2	9	9	ADU70745	Adu70745	Human hep	179	38	62.3	889	8	ADP24012	PRO polyv
107	41	67.2	15	5	ADU71294	Adu71294	Human hep	180	38	62.3	891	2	AAR84338	Maize aco
108	40	65.6	297	5	ABBS4966	Abbs4966	Lactococc	181	38	62.3	891	4	AAU29353	Novel mar
109	40	65.6	297	8	ADS29432	Ads29432	Bacterial	182	38	62.3	891	4	ADS42664	Bacterial
110	40	65.6	532	4	AAG92514	Aag92514	C glutami	183	38	62.3	891	8	ADN18158	Bacterial
111	40	65.6	881	5	ABP28198	Abp28198	Streptoco	184	38	62.3	891	9	AEB40161	L. pneumo
112	40	65.6	881	8	ADV89585	Adv89585	Streptoco	185	38	62.3	891	9	AEB36786	L. pneumo
113	40	65.6	881	8	ADV82992	Adv82992	Streptoco	186	38	62.3	892	6	ABM67426	Phototrab
114	40	65.6	881	8	ADV80838	Adv80838	Streptoco	187	38	62.3	893	8	ADN22413	Bacterial
115	40	65.6	960	6	AEC22866	Aec22866	Protein e	188	38	62.3	895	7	ABO61376	Klebsiell
116	39	63.9	299	9	AEC21206	Aec21206	L. acidop	189	38	62.3	896	8	ADS23030	Bacterial
117	39	63.9	301	9	AEC21150	Aec21150	L. acidop	190	38	62.3	896	8	ADN17556	Bacterial
118	39	63.9	378	7	ABO631658	Abo631658	Klebsiell	191	38	62.3	896	8	ADN21724	Bacterial
119	39	63.9	380	2	AAV17085	Aay17085	Rat hepar	192	38	62.3	896	8	ADN24478	Bacterial
120	39	63.9	536	5	ABBO7812	Abo07812	Rat hepar	193	38	62.3	897	8	ADF54681	Aconitase
121	39	63.9	536	8	ADR88209	Adr88209	Rat hepar	194	38	62.3	898	8	ADQ26200	Arabidops
122	39	63.9	536	8	ADT78176	Adt78176	Rat hepar	195	38	62.3	899	4	ABBB62277	Drosophil
123	39	63.9	536	9	ADY27035	Ady27035	Rat hepar	196	38	62.3	899	5	ABP66115	Bifidobac
124	39	63.9	536	9	AEA42425	Aea42425	Rat hepar	197	38	62.3	900	5	ABB48185	Listeria
125	39	63.9	2931	4	ABBB68229	Abbb68229	Drosophil	198	38	62.3	901	4	ABG81624	S. epider
126	38	62.3	74	3	AAG01360	Aag01360	Human sec	199	38	62.3	901	4	AAG83182	Coryneb
127	38	62.3	90	4	AAW25264	Aaw25264	Human pro	200	38	62.3	901	6	ABM72914	Staphyloc
128	38	62.3	256	4	ABBB64501	Abbb64501	Drosophil	201	38	62.3	901	8	ADN242317	Bacterial
129	38	62.3	380	2	AAV17084	Aay17084	Mouse hep	202	38	62.3	901	8	ADN26748	Bacterial
130	38	62.3	384	5	ABP26811	Abp26811	Streptoco	203	38	62.3	901	8	ADN26679	Bacterial
131	38	62.3	384	8	ADV88250	Adv88250	Streptoco	204	38	62.3	902	4	ABBB61080	Drosophil
132	38	62.3	384	8	ADV79503	Adv79503	Streptoco	205	38	62.3	903	6	ABR64210	Anglogene
133	38	62.3	384	8	ADV81674	Adv81674	Streptoco	206	38	62.3	903	8	ADS28047	Bacterial
134	38	62.3	462	8	ADX88784	Adx88784	Plant ful	207	38	62.3	903	8	ADU06553	Novel bro
135	38	62.3	487	8	ADX73911	Adx73911	Plant ful	208	38	62.3	903	9	ADY70686	Human BAC
136	38	62.3	517	6	ADA89649	Ada89649	Staphyloc	209	38	62.3	904	7	ADF04200	Bacterial
137	38	62.3	535	3	AABO8851	Aab08851	A murine	210	38	62.3	904	8	ADF54679	Aconitase
138	38	62.3	535	5	ABBO7811	Abo07811	Mouse hep	211	38	62.3	906	8	ADN17965	Bacterial
139	38	62.3	535	7	ADG88834	Adg88834	Mouse hpa	212	38	62.3	906	8	ADY24027	Plant ful
140	38	62.3	535	8	ADL16413	Adl16413	Mouse hep	213	38	62.3	908	8	ADS28475	Bacterial
141	38	62.3	535	8	ADM48750	Adm48750	Mouse hpa	214	38	62.3	909	8	ADO59731	B. subtil
142	38	62.3	535	8	ADR88208	Adr88208	Mouse hep	215	38	62.3	909	8	ADS44720	Bacterial
143	38	62.3	535	8	ADT78175	Adt78175	Mouse hep	216	38	62.3	911	8	ADS24683	Bacterial
144	38	62.3	535	9	ADY27033	Ady27033	Murine he	217	38	62.3	914	5	ABP40363	Staphyloc
145	38	62.3	535	9	AEA42424	Aea42424	Mouse hep	218	38	62.3	914	8	ADS04961	Staphyloc
146	38	62.3	553	4	AAV79520	Aab79520	Coryneb	219	38	62.3	914	8	ADS26445	Bacterial
147	38	62.3	557	4	AAV79519	Aab79519	Coryneb	220	38	62.3	915	8	ADN25706	Bacterial
148	38	62.3	573	8	ADX71355	Adx71355	Plant ful	221	38	62.3	915	8	ADS26815	Bacterial
149	38	62.3	631	8	ADX73304	Adx73304	Plant ful	222	38	62.3	919	2	AAR84337	Arabidops
150	38	62.3	643	8	ADX74277	Adx74277	Plant ful	223	38	62.3	922	2	ADS27347	Bacterial
151	38	62.3	655	8	ADX71025	Adx71025	Plant ful	224	38	62.3	927	6	AAR85598	Arabidops
152	38	62.3	668	5	AAW49526	Aam49526	B. mori t	225	38	62.3	927	6	ADA36635	Acinetoba
153	38	62.3	680	8	ADX77402	Adx77402	Plant ful	226	38	62.3	929	8	ADS22231	Bacterial
154	38	62.3	732	3	AAW36178	Aab36178	Mouse sev	227	38	62.3	934	9	ADY60907	Abiotic s
155	38	62.3	732	8	ADO28908	Abo28908	Mouse nov	228	38	62.3	941	8	ADS30597	Bacterial
156	38	62.3	797	6	ABBM5102	Abm65102	Propionib	229	38	62.3	943	8	AAG91445	C. glutami
157	38	62.3	805	4	AAU62827	Aau62827	Propionib	230	38	62.3	943	7	ADD13587	C. glutam
158	38	62.3	805	6	ABBM59346	Abm59346	Propionib	231	38	62.3	943	9	AEB13175	C. glutam
159	38	62.3	824	8	ADX75932	Adx75932	Plant ful	232	38	62.3	968	8	ADX75970	Plant ful
160	38	62.3	848	5	ABBS3982	Abbs3982	Lactococc	233	38	62.3	981	7	ABO68052	Pseudomon
161	38	62.3	848	8	ADS29320	Ads29320	Bacterial	234	38	62.3	1053	6	ABM70125	Phototrab
162	38	62.3	852	2	AAV30948	Aay30948	Human E3	235	38	62.3	1298	9	ADY61065	Abiotic s
163	38	62.3	854	2	AAV30949	Aay30949	Murine E3	236	37	60.7	9	ADU70496	Human hep	
164	38	62.3	854	9	AEA18809	Aea18809	Amino aci	237	37	60.7	15	ADU70883	Human hep	
165	38	62.3	855	9	AED00545	Aed00545	Sulfolobu	238	37	60.7	74	4	ABG25267	Novel hum
166	38	62.3	858	9	ADN26062	Adn26062	Bacterial	239	37	60.7	75	6	ABO00942	Polypepti
167	38	62.3	862	9	AEA18808	Aea18808	Amino aci	240	37	60.7	84	7	ABO67476	Klebsiell
168	38	62.3	869	8	ADS29065	Ads29065	Bacterial	241	37	60.7	97	4	AAU39606	Propionib
169	38	62.3	870	8	ADS44479	Ads44479	Bacterial	242	37	60.7	97	6	ABM36125	Propionib

243	37	60.7	109	2	AAW93582	Aaw93582 Rat rAPO4	316	35	57.4	297	3	AAG56430	Aag56430 Arabidops
244	37	60.7	197	6	ADA35040	Ada35040 Acinetobact	317	35	57.4	317	3	AAG60489	Aag60489 Arabidops
245	37	60.7	198	8	ADX67422	Adx67422 Plant ful	318	35	57.4	317	3	AAG56429	Aag56429 Arabidops
246	37	60.7	384	2	AAI24914	Aai24914 Eisenia f	319	35	57.4	354	3	AB442527	Ab442527 Human ORF
247	37	60.7	730	3	AAAB36179	Aab36179 Rat seven	320	35	57.4	423	5	ABP73924	Abp73924 Candida a
248	37	60.7	756	6	AAE330471	Aae330471 Haemophil	321	35	57.4	423	6	ABR70164	Ab70164 C. neofo
249	37	60.7	784	7	ADZ31168	Adz31168 Human dia	322	35	57.4	428	6	ABR53385	Ab53385 Protein s
250	37	60.7	805	8	ADN23021	Adn23021 Bacterial	323	35	57.4	428	7	ADK63326	Adk63326 Disease t
251	37	60.7	805	8	ADN23020	Adn23020 Bacterial	324	35	57.4	487	3	AAG28042	Aag28042 Arabidops
252	37	60.7	1192	4	ABG02038	Abg02038 Novel hum	325	35	57.4	540	5	AAM49664	Aam49664 Rat SmGlu
253	37	60.7	1653	7	ADF74137	Adf74137 Human nov	326	35	57.4	554	3	AB20994	Ab20994 Human Rec
254	37	60.7	2113	4	ABB64885	Abb64885 Drosophil	327	35	57.4	564	2	AAM48788	Aam48788 Thyroid p
255	37	60.7	2113	8	ADR96394	Adr96394 Drosophil	328	35	57.4	569	3	AAG28041	Aag28041 Arabidops
256	36.5	59.8	143	6	ABR42559	Ab42559 Novobioci	329	35	57.4	569	3	AAG38775	Aag38775 Arabidops
257	36	59.0	9	ADU70820	Ad70820 Human hep	330	35	57.4	621	3	AAG28040	Aag28040 Arabidops	
258	36	59.0	15	ADU71088	Ad71088 Human hep	331	35	57.4	623	3	AAG38774	Aag38774 Arabidops	
259	36	59.0	202	8	ADQ08768	Adq08768 Ciona int	332	35	57.4	624	2	AAM48789	Aam48789 Thyroid p
260	36	59.0	218	8	ADH39781	Adh39781 Streptomy	333	35	57.4	654	7	ADC31447	Adc31447 Human nov
261	36	59.0	252	3	AAG28249	Aag28249 Arabidops	334	35	57.4	681	2	AAM48786	Aam48786 Thyroid p
262	36	59.0	252	3	AAG46618	Aag46618 Arabidops	335	35	57.4	689	2	AAM48787	Aam48787 Thyroid p
263	36	59.0	262	3	AAG46617	Aag46617 Arabidops	336	35	57.4	689	8	ADS23685	Ads23685 Bacteri
264	36	59.0	262	3	AAG28248	Aag28248 Arabidops	337	35	57.4	701	6	ADA34430	Ada34430 Acinetoba
265	36	59.0	279	7	ADF04263	Adf04263 Bacterial	338	35	57.4	731	3	AAG38773	Aag38773 Arabidops
266	36	59.0	341	3	AAG46616	Aag46616 Arabidops	339	35	57.4	740	2	AAM48790	Aam48790 Thyroid p
267	36	59.0	341	3	AAG28247	Aag28247 Arabidops	340	35	57.4	752	8	ADX95153	Adx95153 Plant ful
268	36	59.0	341	5	AB93596	Ab93596 Herbicida	341	35	57.4	756	3	AB20995	Ab20995 Human Rec
269	36	59.0	463	3	AAI58263	Aai58263 Guar gala	342	35	57.4	784	2	AAM48783	Aam48783 Thyroid p
270	36	59.0	470	3	AAI05940	Aai05940 Protein d	343	35	57.4	785	8	ABM83658	Abm83658 Human dia
271	36	59.0	481	2	AAI72095	Aai72095 Human mGL	344	35	57.4	813	6	ABO52993	Ab052993 Human put
272	36	59.0	523	5	ABB07814	Abb07814 Chicken h	345	35	57.4	842	8	ABM83657	Abm83657 Human dia
273	36	59.0	523	8	ABW02017	Abw02017 Chicken h	346	35	57.4	848	2	AAI07733	Aai07733 Human thy
274	36	59.0	523	8	ADR88211	Adr88211 Chicken h	347	35	57.4	852	2	AAM48782	Aam48782 Thyroid p
275	36	59.0	523	9	ADY78178	Ady78178 Chicken h	348	35	57.4	868	6	ABU35640	Abu35640 Haemophil
276	36	59.0	523	9	ADY27037	Ady27037 Chicken h	349	35	57.4	868	6	ABU30578	Abu30578 Protein e
277	36	59.0	523	9	AEA42427	Aea42427 Chicken h	350	35	57.4	881	2	AAM48791	Aam48791 Thyroid p
278	36	59.0	577	8	ADM48198	Adm48198 Polypepti	351	35	57.4	888	7	AD47732	Ad47732 Human NOV
279	36	59.0	594	8	ADO07212	Ado07212 Streptoco	352	35	57.4	888	8	ADJ79002	Adj79002 Human NOV
280	36	59.0	657	8	ADO07196	Ado07196 Streptoco	353	35	57.4	915	2	AAR80479	Aar80479 Rat metab
281	36	59.0	659	7	ADF04463	Adf04463 Bacterial	354	35	57.4	915	6	AAE30199	Aae30199 Rat metab
282	36	59.0	728	4	AAG98830	E. coli g	355	35	57.4	915	7	ADD48384	Add48384 Rat Prote
283	36	59.0	728	6	ABU14683	Abu14683 Protein e	356	35	57.4	915	7	ADE55967	Ad555967 Rat Prote
284	36	59.0	867	2	AAU14683	Aau14683 Human mGL	357	35	57.4	915	8	ADO29104	Ado29104 Mouse nov
285	36	59.0	898	9	ADZ67613	Adz67613 Chimeric	358	35	57.4	924	2	AAR35445	Aar35445 Human TPO
286	36	59.0	907	9	ADZ67611	Adz67611 Chimeric	359	35	57.4	933	2	AAR35445	Aar35445 Human TPO
287	36	59.0	910	9	ADZ67609	Adz67609 Chimeric	360	35	57.4	933	2	AAR32875	Aar32875 Human TPO
288	36	59.0	912	9	ADZ67607	Adz67607 Chimeric	361	35	57.4	933	2	AAR75689	Aar75689 Human thr
289	36	59.0	913	10	AEF72174	Aef72174 Human tar	362	35	57.4	933	7	ADJ68762	Adj68762 Human hea
290	36	59.0	915	2	AAR72097	Aar72097 Human mGL	363	35	57.4	933	8	ADQ14321	Adq14321 Human thy
291	36	59.0	915	5	ABG95165	Abg95165 Human GPC	364	35	57.4	933	8	ADR14768	Adr14768 Human thy
292	36	59.0	915	5	ABG95155	Abg95155 Human GPC	365	35	57.4	933	8	ADR14361	Adr14361 Human thy
293	36	59.0	915	5	ABG95164	Abg95164 Human GPC	366	35	57.4	933	9	ABE77781	Aeb77781 Human thy
294	36	59.0	915	5	ABG95166	Abg95166 Human GPC	367	35	57.4	933	10	AEF07493	Aef07493 Human thy
295	36	59.0	915	5	ABG95163	Abg95163 Human GPC	368	35	57.4	948	2	AAM48781	Aam48781 Thyroid p
296	36	59.0	915	6	ABP81849	Abp81849 Human met	369	35	57.4	1142	8	ABM84821	Abm84821 Human dia
297	36	59.0	915	7	ADJ55969	Adj55969 Human Pro	370	35	57.4	1144	7	ADD01202	Add01202 Human hel
298	36	59.0	915	7	ADJ93191	Adj93191 Human met	371	35	57.4	1208	2	AAW95050	Aaw95050 Human hel
299	36	59.0	915	8	ADO29103	Ado29103 Human nov	372	35	57.4	1208	3	AAW95050	Aaw95050 Human hel
300	36	59.0	915	9	ADZ67603	Adz67603 Human met	373	35	57.4	1208	5	ABG93387	Abg93387 Human Rec
301	36	59.0	915	9	ABE87480	Ab87480 Human met	374	35	57.4	1208	8	ADQ21562	Adq21562 Human sof
302	36	59.0	915	10	AEF72172	Aef72172 Human tar	375	35	57.4	1208	8	ADQ21562	Adq21562 Human sof
303	36	59.0	922	2	AAR72098	Aar72098 Human mGL	376	35	57.4	1215	6	ABM15836	Abm15836 Mycobacte
304	36	59.0	922	10	AEF72173	Aef72173 Human tar	377	35	57.4	1291	4	ABG28604	Abg28604 Novel hum
305	36	59.0	1028	7	ADF28592	Adf28592 Murine PA	378	35	57.4	1306	8	ADM90956	Adm90956 Human pha
306	35	57.4	11	2	AAI17080	Aai17080 Human pla	379	35	57.4	3913	6	ABM67350	Abm67350 Photorhab
307	35	57.4	68	4	AAU63347	Aau63347 Propionib	380	35	57.4	10182	5	ABP38314	Abp38314 Staphyloc
308	35	57.4	68	6	ABM59866	Abm59866 Propionib	381	34	55.7	17	9	ADP26216	Adp26216 Human EA4
309	35	57.4	88	9	AEC37224	Aec37224 ReIE-toxi	382	34	55.7	17	9	AEA17140	Aea17140 Human TNF
310	35	57.4	120	5	ABP02418	Abp02418 Human ORF	383	34	55.7	17	9	ABE17203	Aeb17203 Epha2-spe
311	35	57.4	185	3	AAG56431	Aag56431 Arabidops	384	34	55.7	17	9	AEA43047	Aea43047 Epha2 ant
312	35	57.4	185	3	AAG60491	Aag60491 Arabidops	385	34	55.7	19	10	AEF39151	Aef39151 Human ser
313	35	57.4	213	5	ABP07167	Abp07167 Human ORF	386	34	55.7	22	10	AEF06087	Aef06087 Human hea
314	35	57.4	285	8	ADN47107	Adn47107 Thermoco	387	34	55.7	71	5	AAU98096	Aau98096 Human bet
315	35	57.4	297	3	AAG60490	Aag60490 Arabidops	388	34	55.7	78	7	ADH86762	Adh86762 Enterococ

389	34	55.7	98	4	AAB80990	Human ant	462	34	55.7	937	4	AAU02144	Aau02144 Rx 28, mo
390	34	55.7	98	5	ABG78169	Human Fv	463	34	55.7	937	4	AAU02145	Aau02145 Rx 72, mo
391	34	55.7	98	5	ABG78168	Human Fv	464	34	55.7	937	4	AAU02147	Aau02147 Rx 193, m
392	34	55.7	98	5	ABG91860	Human ant	465	34	55.7	937	4	AAU02146	Aau02146 Rx 39, mo
393	34	55.7	98	5	ABG91859	Human ant	466	34	55.7	937	4	AAU02143	Aau02143 Rx 25, mo
394	34	55.7	98	6	ABO27114	Human ger	467	34	55.7	937	4	AAU02148	Aau02148 Rx 7, mod
395	34	55.7	98	7	ADF10159	Antibody	468	34	55.7	938	3	AAU45004	Protein e
396	34	55.7	98	7	ADF09951	Antibody	469	34	55.7	1402	8	ADN22849	Bacterial
397	34	55.7	98	7	ADF10056	VEGF anti	470	34	55.7	2539	8	ADU25430	L. acidop
398	34	55.7	98	7	ADJ80331	VF gene 1	471	33.5	54.9	674	4	ABB61427	Drosophil
399	34	55.7	98	9	ADY75336	Protein e	472	33.5	54.9	1002	3	AAI79166	Pneumocys
400	34	55.7	98	9	ABE13612	Human var	473	33	54.1	44	5	AAE21953	Human lmm
401	34	55.7	98	9	AEC20851	Human var	474	33	54.1	49	2	AAW28322	Staphyloc
402	34	55.7	98	9	AEC36757	Human ger	475	33	54.1	53	5	ABP04205	Human ORF
403	34	55.7	98	10	AE847570	Human CDR	476	33	54.1	55	4	ABB50601	Human sec
404	34	55.7	100	10	AEF81801	Human ant	477	33	54.1	55	6	ABO44858	Novel hum
405	34	55.7	109	9	ADU26551	Human ant	478	33	54.1	55	7	ABO26338	Protein a
406	34	55.7	117	9	AEA17137	Human TNF	479	33	54.1	57	5	ABP08149	Human ORF
407	34	55.7	119	7	ABR61523	Human RF-	480	33	54.1	63	4	AAU64687	Propionib
408	34	55.7	120	2	AAR65161	Human bcl	481	33	54.1	63	6	ABM61206	Propionib
409	34	55.7	120	7	ABR83200	Human ant	482	33	54.1	92	4	AAI91994	Mutant hu
410	34	55.7	123	9	ADW26196	Human EA4	483	33	54.1	93	4	AAO10801	Human pol
411	34	55.7	123	9	AEI17201	EphA2-spe	484	33	54.1	100	4	ABB50599	Human sec
412	34	55.7	123	9	AEA43045	EphA2 ant	485	33	54.1	100	6	ABO44856	Novel hum
413	34	55.7	126	2	AAI21147	Human bcl	486	33	54.1	100	7	ABO26336	Protein a
414	34	55.7	136	5	ABB92774	Herbicida	487	33	54.1	102	4	AAO12801	Human pol
415	34	55.7	146	4	AB203318	Wheat apo	488	33	54.1	124	4	ABB60176	Drosophil
416	34	55.7	149	6	ABJ36935	Anti-CD40	489	33	54.1	130	8	ADX75159	Plant ful
417	34	55.7	152	9	AEC81264	Human mon	490	33	54.1	132	3	AAG30247	Arabidops
418	34	55.7	160	7	ADE09227	Novel pro	491	33	54.1	133	4	AAU62616	Propionib
419	34	55.7	200	9	ABE17262	EphA4-spe	492	33	54.1	133	6	ABM59135	Propionib
420	34	55.7	233	7	ADF07924	Bacterial	493	33	54.1	164	3	AAG30246	Arabidops
421	34	55.7	238	9	ADW26212	Human scF	494	33	54.1	178	8	ADY05698	Plant ful
422	34	55.7	239	2	AAW46962	Human scF	495	33	54.1	181	10	AEF65943	B. subtil
423	34	55.7	259	9	ADW26210	Human scF	496	33	54.1	185	4	AAU35490	Haemophil
424	34	55.7	260	6	ABU19436	Protein e	497	33	54.1	185	6	ABU30345	Protein e
425	34	55.7	261	6	ABU21635	Protein e	498	33	54.1	201	4	ABB65657	Drosophil
426	34	55.7	262	6	ABU22621	Protein e	499	33	54.1	202	3	AAG30245	Arabidops
427	34	55.7	274	6	ABM70789	Staphyloc	500	33	54.1	209	8	ADX73581	Plant ful
428	34	55.7	279	9	ADW26208	Human scF	501	33	54.1	213	7	ABM73698	DNA clone
429	34	55.7	289	5	AB53498	Lactococc	502	33	54.1	213	8	ADY05303	Plant ful
430	34	55.7	289	8	ADS29274	Bacterial	503	33	54.1	224	4	ABM67474	Amino aci
431	34	55.7	339	9	AEA43058	Single ch	504	33	54.1	224	8	ADU04688	Mumps vir
432	34	55.7	373	6	ABM70683	Photorhab	505	33	54.1	225	8	ADT87034	Yeast Str
433	34	55.7	396	4	AAU02152	Plant res	506	33	54.1	263	8	ADX78221	Plant ful
434	34	55.7	430	4	ABE68638	Drosophil	507	33	54.1	263	8	ADY07670	Plant ful
435	34	55.7	441	4	ABG02026	Novel hum	508	33	54.1	269	9	AEBA2538	L. pneumo
436	34	55.7	449	3	ABG22958	Arabidops	509	33	54.1	275	6	ABU94235	Trifolium
437	34	55.7	455	4	ABB71816	Drosophil	510	33	54.1	276	2	AAI37171	Amino aci
438	34	55.7	467	3	AAG40087	Arabidops	511	33	54.1	306	3	AAG09000	Arabidops
439	34	55.7	467	3	AAG41408	Arabidops	512	33	54.1	306	3	AAG42634	Arabidops
440	34	55.7	467	3	AAG22957	Arabidops	513	33	54.1	325	2	AAW97705	Staphyloc
441	34	55.7	485	2	AAW01459	Arabidops	514	33	54.1	330	3	ABM16019	E. coli p
442	34	55.7	485	3	AAG40086	Arabidops	515	33	54.1	330	4	AAG98321	Escherich
443	34	55.7	485	3	AAG41407	Arabidops	516	33	54.1	334	4	AAU19853	Human nov
444	34	55.7	485	7	ADF75199	Thale cre	517	33	54.1	334	4	AAU87678	Novel cen
445	34	55.7	485	8	ADN72983	Thale cre	518	33	54.1	334	5	ABP48073	Human pol
446	34	55.7	489	3	AAG22956	Arabidops	519	33	54.1	334	7	ADC11035	Human pro
447	34	55.7	497	3	AAG41406	Arabidops	520	33	54.1	334	8	ADI54993	Novel hum
448	34	55.7	497	7	ADF06409	Bacterial	521	33	54.1	342	8	ADP04517	Sea squir
449	34	55.7	508	3	AAG40085	Arabidops	522	33	54.1	363	3	AAI77125	Human neu
450	34	55.7	510	8	ADN21556	Bacterial	523	33	54.1	387	4	ABO94592	Human pro
451	34	55.7	578	2	AA81844	Human afa	524	33	54.1	388	7	ABO74233	Pseudomon
452	34	55.7	599	2	AA81845	Human afa	525	33	54.1	389	9	ADM05762	Human pro
453	34	55.7	599	2	AA81845	Human afa	526	33	54.1	389	9	AEC88692	Human CDN
454	34	55.7	634	6	ABM67337	Photorhab	527	33	54.1	391	6	ABU94276	Trifolium
455	34	55.7	643	8	ADL91024	Rat manno	528	33	54.1	391	8	ADU03908	White clo
456	34	55.7	669	6	ABU44839	Protein e	529	33	54.1	399	7	ADF05840	Bacterial
457	34	55.7	679	9	AED53582	Trypanoso	530	33	54.1	410	8	ADR10318	Human pro
458	34	55.7	736	5	ABP65381	Bifidobac	531	33	54.1	423	8	ADQ29657	Human col
459	34	55.7	751	8	ADK16547	Nanoarcha	532	33	54.1	423	8	ADP24748	PRO polyP
460	34	55.7	912	3	AAI44818	Potato Gp	533	33	54.1	423	8	ADP24748	PRO polyP
461	34	55.7	937	2	AAI52152	Potato re	534	33	54.1	423	9	AEA04485	Human pro

535	33	54.1	423	10	Aef69969	Aef69969 Colorecta	608	33	54.1	3144	2	AAW36887	AAW36887 Previousl
536	33	54.1	428	5	ABB90572	Abb90572 Chlamydia	609	33	54.1	3144	2	AAW09871	Aaw09871 Human hun
537	33	54.1	428	6	ABU27068	Abu27068 Protein e	610	33	54.1	3144	2	AAW44742	Aaw44742 Human hun
538	33	54.1	428	9	AC935688	Aec95688 C. pneumo	611	33	54.1	3144	2	AY334493	Ady33493 Human hun
539	33	54.1	452	2	ACY35435	Aay35435 C. pneumo	612	33	54.1	3144	9	ADY98141	Ady98141 Human hun
540	33	54.1	453	8	ADR95252	Aes95252 Novel S.	613	33	54.1	3144	9	AEC36191	Aec36191 Anino aci
541	33	54.1	453	9	AEA59122	Aes59122 Streptoco	614	33	54.1	3223	4	ABE11407	Abb11407 Human Hun
542	33	54.1	460	7	ADP06963	Adf06963 Bacterial	615	33	54.1	3223	4	ABB11470	Abb11470 Human Hun
543	33	54.1	482	2	ARF75382	Aar75382 Natural r	616	33	54.1	3572	5	ABG95659	Abg95659 Human nuc
544	33	54.1	482	2	ARF75386	Aar75386 Natural r	617	33	54.1	4970	9	ABE00354	Aeb00354 Plastid c
545	33	54.1	488	8	ADT59749	Aut59749 Plant pol	618	32.5	53.3	537	7	ABO81643	Abos81643 Pseudomon
546	33	54.1	500	4	AAU19735	Aau19735 Human nov	619	32.5	53.3	2359	6	ABP56959	Abp56959 E. maxima
547	33	54.1	500	5	ABP47955	Abp47955 Human pol	620	32.5	53.3	2360	6	ABP56972	Abp56972 E. maxima
548	33	54.1	500	7	ADC10917	Adc10917 Human ext	621	32.5	53.3	2661	6	ABP56958	Abp56958 Bimeria m
549	33	54.1	532	6	ABP77145	Abp77145 N. gonorr	622	32	52.5	12	9	ABE13404	Aeb13404 Tag #5 su
550	33	54.1	532	10	AEF72317	Aee72317 Human tar	623	32	52.5	13	5	AAW48823	Aam48823 Endostat1
551	33	54.1	533	4	AY511175	Aay511175 Human bra	624	32	52.5	14	3	AAW35855	Aam35855 T7 phage
552	33	54.1	533	4	AAU29021	Aau29021 Human PAR	625	32	52.5	14	9	ADV58994	Adv58994 G protein
553	33	54.1	533	10	AE86985	Aee86985 Human che	626	32	52.5	14	9	ADV58441	Adv58441 G protein
554	33	54.1	535	6	ABU00691	Abu00691 S. pneumo	627	32	52.5	15	9	ADV58450	Adv58450 G protein
555	33	54.1	536	8	ADK47489	Adk47489 Streptoco	628	32	52.5	15	9	ADV59027	Adv59027 G protein
556	33	54.1	540	3	AAYS1176	Aay51176 Human ute	629	32	52.5	17	3	AAI49353	Aay49353 P. aerugi
557	33	54.1	543	2	AAW70466	Aaw70466 South Afr	630	32	52.5	17	3	AAI49354	Aay49354 P. aerugi
558	33	54.1	548	4	AAU87401	Aau87401 Novel Cen	631	32	52.5	17	8	ADG73765	Adg73765 P. aerugi
559	33	54.1	548	8	ADI54716	Adi54716 Novel hum	632	32	52.5	17	8	ADG73764	Adg73764 P. aerugi
560	33	54.1	550	8	ADS29184	Ads29184 Bacterial	633	32	52.5	18	2	AAI42663	Aay42663 HHV-7 pep
561	33	54.1	560	2	AAW70473	Aaw70473 Girwood	634	32	52.5	44	8	ABO54680	Abos4680 Human gen
562	33	54.1	561	8	ADRI14159	Adri14159 Human NF-	635	32	52.5	53	4	ABBI16824	Abbi16824 Human ner
563	33	54.1	561	8	ADP23478	Adp23478 PRO polyp	636	32	52.5	56	7	ADC27504	Adc27504 Western e
564	33	54.1	561	9	ADV73219	Adv73219 Human col	637	32	52.5	59	2	AAW04957	Aaw04957 Partial F
565	33	54.1	565	4	ABG23778	Abg23778 Novel hum	638	32	52.5	59	9	AED55900	Aed55900 Fusarium
566	33	54.1	568	7	ADD22422	Add22422 HLA-B46 T	639	32	52.5	80	3	AB587733	Ab587733 Breast an
567	33	54.1	568	7	ADD45724	Add45724 Human Pro	640	32	52.5	85	6	ABP80108	Abp80108 N. gonorr
568	33	54.1	568	7	ADE54650	Ade54650 Human Pro	641	32	52.5	92	4	AAW51744	Aam51744 Human FSH
569	33	54.1	568	7	ADE54646	Ade54646 Human Pro	642	32	52.5	93	7	ABO83220	Abos83220 Pseudomon
570	33	54.1	568	7	ADI15912	Adi15912 Human pp	643	32	52.5	99	2	AAW95288	Aaw95288 Chlamydia
571	33	54.1	568	8	ADO08084	Ado08084 Human pol	644	32	52.5	100	2	AAW95284	Aaw95284 Chlamydia
572	33	54.1	568	8	AEC12341	Aec12341 Human sur	645	32	52.5	105	7	ADD25141	Add25141 Fertility
573	33	54.1	568	9	AEC12726	Aec12726 Human sur	646	32	52.5	105	8	ADN61156	Adn61156 Radish nu
574	33	54.1	582	5	ABP73516	Abp73516 Candida a	647	32	52.5	117	2	AAI37327	Aay37327 Amino aci
575	33	54.1	583	8	ADN23432	Adn23432 Bacterial	648	32	52.5	117	4	ADM20025	Adm20025 Protein e
576	33	54.1	610	5	ABG61879	Abg61879 Prostate	649	32	52.5	121	4	AAU54626	Aau54626 Propionib
577	33	54.1	610	5	ADB75244	Adb75244 Prostate	650	32	52.5	121	6	ABM51145	Abm51145 Propionib
578	33	54.1	610	8	ADLI3218	Adli3218 Human ste	651	32	52.5	124	4	ADM19765	Adm19765 Protein e
579	33	54.1	611	8	ADT66701	Adt66701 Murine ca	652	32	52.5	125	3	AB40264	Ab40264 Human ORF
580	33	54.1	611	10	AEF79055	Aee79055 Spatzle 3	653	32	52.5	125	5	ABP34931	Abp34931 Human ORF
581	33	54.1	613	5	AAE22074	Aae22074 Gastrero	654	32	52.5	134	2	AAW95286	Aaw95286 Chlamydia
582	33	54.1	614	6	ABM72873	Abm72873 Staphyloc	655	32	52.5	145	2	AAW15094	Aaw15094 hCG/hTSH
583	33	54.1	617	8	ADR66942	Adr66942 Human pro	656	32	52.5	149	7	ABO83524	Abos83524 Pseudomon
584	33	54.1	617	8	ADR66044	Adr66044 Human pro	657	32	52.5	154	6	ABU08012	Abu08012 Soybean m
585	33	54.1	637	4	ABG23529	Abg23529 Novel hum	658	32	52.5	158	6	ABU08009	Abu08009 Soybean m
586	33	54.1	641	9	ADY70667	Ady70667 Human nic	659	32	52.5	179	8	ADS28182	Ads28182 Bacterial
587	33	54.1	644	8	ADY23623	Ady23623 Bacterial	660	32	52.5	189	3	AG24503	Aag24503 Arachidops
588	33	54.1	728	7	ADB70233	Adb70233 C. neofo	661	32	52.5	206	8	ADU66592	Adu66592 Human kin
589	33	54.1	803	8	ADJ49938	Adj49938 Oil-asso	662	32	52.5	210	7	ADF05490	Adf05490 Bacterial
590	33	54.1	831	8	ADS21941	Ads21941 Bacterial	663	32	52.5	221	7	ABO71523	Abos71523 Pseudomon
591	33	54.1	869	4	ABW57877	Abw57877 Drosophil	664	32	52.5	226	4	ABG24185	Abg24185 Novel hum
592	33	54.1	870	5	ADJ49939	Adj49939 Oil-asso	665	32	52.5	234	8	ADY06333	Ady06333 Plant ful
593	33	54.1	910	5	AAE22072	Aae22072 Gastrero	666	32	52.5	240	2	AAW52309	Aaw52309 Modified
594	33	54.1	951	5	ABP61509	Abp61509 Human NF-	667	32	52.5	244	1	AAW70561	Aap70561 Product o
595	33	54.1	985	4	ABM60399	Abm60399 Drosophil	668	32	52.5	257	9	ADV09395	Adv09395 TFLCSV ca
596	33	54.1	1037	9	ABE91540	Aeb91540 Microbial	669	32	52.5	261	1	AAW70406	Aap70406 ORF 3 gen
597	33	54.1	1050	5	ABM06245	Abm06245 Bacillus	670	32	52.5	263	8	ADN22777	Adn22777 Bacterial
598	33	54.1	1142	4	ABM62802	Abm62802 Drosophil	671	32	52.5	274	4	AAW93889	Aam93889 Human pol
599	33	54.1	1142	8	ADO08088	Ado08088 Fly polyp	672	32	52.5	274	8	ADLI31986	Adli31986 Human pro
600	33	54.1	1191	5	AAU98061	Aau98061 Bacillus	673	32	52.5	277	3	AAI82393	Aay82393 C. tracho
601	33	54.1	1347	8	ADN72391	Adn72391 Thale cre	674	32	52.5	290	4	AAU30224	Aau30224 Novel hum
602	33	54.1	1401	8	ADLI3306	Adli3306 Human ste	675	32	52.5	290	7	ABM86508	Abm86508 Rice abio
603	33	54.1	1401	9	ADX07508	Adx07508 Cyclin-de	676	32	52.5	311	5	AAU85734	Aau85734 Human G-P
604	33	54.1	1887	6	ABU20638	Abu20638 Protein e	677	32	52.5	311	5	AAU85735	Aau85735 Human G-P
605	33	54.1	3139	10	AE339884	Aee33984 Amino aci	678	32	52.5	312	4	AAU71972	Aag71972 Human olf
606	33	54.1	3141	7	ADJ70444	Adj70444 Human hea	679	32	52.5	312	4	AAU24711	Aau24711 Human olf
607	33	54.1	3144	2	AAW58777	Aar58777 Protein e	680	32	52.5	312	5	ABP95827	Abp95827 Human GPC

681	32	52.5	312	5	AAU95659	Aau95659 Human olf	754	32	52.5	538	5	AAE23297	Aae23297 Human nec
682	32	52.5	312	5	AAU95776	Aau95776 Human olf	755	32	52.5	538	6	ABJ20238	Abj20238 Human IG
683	32	52.5	312	5	AAU95331	Aau95331 G-coupled	756	32	52.5	538	7	ADF13699	Adf13699 Tumor-Aas
684	32	52.5	312	7	ADC85853	Adc85853 Human GPC	757	32	52.5	538	8	ADE86687	Ade86687 Human poi
685	32	52.5	312	7	ADC85621	Adc85621 Human GPC	758	32	52.5	538	9	AED21667	Aed21667 Human can
686	32	52.5	312	8	ADG83522	Adg83522 Human Olf	759	32	52.5	540	8	ADN26066	Adn26066 Bacterial
687	32	52.5	316	6	ADA15561	Ada15561 A. thalia	760	32	52.5	546	8	ABM83047	Abm83047 Human dia
688	32	52.5	316	7	ADC46669	Adc46669 Thalecres	761	32	52.5	555	4	ABG16842	Abg16842 Novel hum
689	32	52.5	316	7	ADG55808	Adg55808 Thalecres	762	32	52.5	560	6	ABU09237	Abu09237 Human G-p
690	32	52.5	316	7	ADD30522	Add30522 Plant yie	763	32	52.5	560	8	ADO28902	Ado28902 Human nov
691	32	52.5	316	8	ADI43955	Adi43955 Plant tra	764	32	52.5	562	2	AAV41268	Aav41268 Protein h
692	32	52.5	318	6	ABR01613	Abr01613 Human G p	765	32	52.5	571	8	ADR09433	Adr09433 Human pro
693	32	52.5	319	9	ABM94945	Abm94945 M. xanthu	766	32	52.5	572	5	ABP29715	Abp29715 Streptoco
694	32	52.5	323	8	ADSA1881	Ads1881 Bacterial	767	32	52.5	572	5	ABP26316	Abp26316 Streptoco
695	32	52.5	325	8	ADNI18591	Adni18591 Bacterial	768	32	52.5	577	7	ADC06791	Adc06791 Human pro
696	32	52.5	333	4	ABBG1558	Abbg1558 Drosophil	769	32	52.5	577	7	ADC06792	Adc06792 Prostate
697	32	52.5	333	4	ABG16843	Abg16843 Novel hum	770	32	52.5	581	3	AAV84322	Aav84322 A human c
698	32	52.5	335	8	ADN23606	Adn23606 Bacterial	771	32	52.5	581	4	ABM65701	Abm65701 Novel pro
699	32	52.5	340	8	ADS24784	Ads24784 Bacterial	772	32	52.5	581	8	ADI29310	Adi29310 Human MAR
700	32	52.5	341	4	AAB96499	Aab96499 Putative	773	32	52.5	581	8	ADQ15154	Adq15154 Human can
701	32	52.5	341	6	ABU42121	Abu42121 Protein e	774	32	52.5	582	4	AAU03896	Aau03896 G protein
702	32	52.5	341	6	ABU39552	Abu39552 Protein e	775	32	52.5	584	9	AEB40592	Aeb40592 L. pneumo
703	32	52.5	345	2	AAW41166	Aaw41166 Metal-reg	776	32	52.5	591	9	ABM97017	Abm97017 M. xanthu
704	32	52.5	345	8	ABT06032	Abt06032 Soil rest	777	32	52.5	595	5	ABP65875	Abp65875 Bifidobac
705	32	52.5	346	4	ABB71526	Abb71526 Drosophil	778	32	52.5	599	9	AEB37268	Aeb37268 L. pneumo
706	32	52.5	349	3	AY82392	AY82392 C. tracho	779	32	52.5	605	2	AAW84086	Aaw84086 Human mem
707	32	52.5	353	8	ADN21396	Adn21396 Bacterial	780	32	52.5	606	2	AAW84086	Aaw84086 Human mem
708	32	52.5	371	6	ABU32109	Abu32109 Protein e	781	32	52.5	606	3	AAV77489	Aav77489 Human WD-
709	32	52.5	373	2	AAW76362	Aaw76362 C. tracho	782	32	52.5	606	7	ADM05157	Adm05157 Human pro
710	32	52.5	375	8	ADN25794	Adn25794 Bacterial	783	32	52.5	606	7	ADM03646	Adm03646 Antipsori
711	32	52.5	387	7	ABO66852	AbO66852 Klebsiell	784	32	52.5	606	8	ABM80560	Abm80560 Tumour-as
712	32	52.5	388	8	ADSA4090	Ads44090 Bacterial	785	32	52.5	606	8	ADT04108	Adt04108 Novel hum
713	32	52.5	393	2	AY56759	AY56759 C. tracho	786	32	52.5	606	9	AEC88087	Aec88087 Human GDN
714	32	52.5	394	1	AAW60004	Aaw60004 Sequence	787	32	52.5	608	8	ADX68346	Adx68346 Plant ful
715	32	52.5	394	2	AAW15149	Aaw15149 Chlamydia	788	32	52.5	627	4	ABG16844	Abg16844 Novel hum
716	32	52.5	394	2	AAW73141	Aaw73141 Chlamydia	789	32	52.5	635	7	ABO60660	AbO60660 Klebsiell
717	32	52.5	394	2	AAW57775	Aaw57775 Chlamydia	790	32	52.5	637	7	ADM04513	Adm04513 Human pro
718	32	52.5	394	2	AY56761	AY56761 C. tracho	791	32	52.5	637	9	AEC87443	Aec87443 Human GDN
719	32	52.5	394	2	AY537494	AY537494 Chlamydia	792	32	52.5	641	6	ABM68514	Abm68514 Photorhab
720	32	52.5	394	3	AY81268	AY81268 Chlamydia	793	32	52.5	645	8	ADX68345	Adx68345 Plant ful
721	32	52.5	395	5	ABBS4242	Abbs4242 Lactococc	794	32	52.5	647	9	AEB87861	Aeb87861 Lipomyces
722	32	52.5	400	5	ABP29950	Abp29950 Streptoco	795	32	52.5	666	8	ADY07866	Ady07866 Plant ful
723	32	52.5	400	8	ADR83914	Adr83914 S. pyogen	796	32	52.5	672	5	ABP73226	Abp73226 Candida a
724	32	52.5	415	7	ABO83146	AbO83146 Pseudomon	797	32	52.5	694	7	ADB64315	Adb64315 Human pro
725	32	52.5	422	8	ADS22293	Ads22293 Bacterial	798	32	52.5	694	7	ADC06790	Adc06790 Human pro
726	32	52.5	423	2	AAW27635	Aaw27635 Streptoco	799	32	52.5	694	8	ADQ66699	Adq66699 Novel hum
727	32	52.5	423	4	AAW15541	Aaw15541 Peptide #	800	32	52.5	730	7	ADP71289	Adp71289 Novel hum
728	32	52.5	436	4	ABW29374	Abw29374 Peptide #	801	32	52.5	741	6	ADA34611	Ada34611 Acinetoba
729	32	52.5	436	4	AAW55334	Aaw55334 Human bra	802	32	52.5	743	8	ADV88179	Adv88179 Streptoco
730	32	52.5	436	4	AAW03290	Aaw03290 Peptide #	803	32	52.5	743	8	ADV79432	Adv79432 Streptoco
731	32	52.5	436	5	ABG37291	Abg37291 Human pep	804	32	52.5	750	5	ADP27691	Adp27691 Streptoco
732	32	52.5	443	4	AAW82383	Aaw82383 S. epider	805	32	52.5	750	3	ADV81607	Adv81607 Streptoco
733	32	52.5	444	5	ABP39501	Abp39501 Staphyloc	806	32	52.5	758	3	AAW11540	Aaw11540 SEN virus
734	32	52.5	444	8	ADSO5410	Adso5410 Staphyloc	807	32	52.5	798	7	ADF70427	Adf70427 Orphan re
735	32	52.5	448	4	ABBG68176	Abbg68176 Drosophil	808	32	52.5	818	9	ADX05054	Adx05054 Cat IL4 r
736	32	52.5	454	3	AY82390	AY82390 C. tracho	809	32	52.5	819	4	ABBG6041	Abbg6041 Drosophil
737	32	52.5	457	3	AAW82388	Aaw82388 C. tracho	810	32	52.5	823	9	ADX05055	Adx05055 Dog IL4 r
738	32	52.5	458	4	AAW25635	Aaw25635 Human pro	811	32	52.5	842	4	ABW65111	Abw65111 Drosophil
739	32	52.5	466	10	AE888032	Aee88032 Human pro	812	32	52.5	882	5	ABP25774	Abp25774 Streptoco
740	32	52.5	471	9	ABW92751	Abw92751 M. xanthu	813	32	52.5	882	5	ABU46739	Abu46739 Protein e
741	32	52.5	474	4	AAW62619	Aaw62619 Petunia z	814	32	52.5	883	6	AAU06828	Aau06828 S. pneumo
742	32	52.5	474	4	AAW64525	Aaw64525 P. hybrid	815	32	52.5	883	2	AAU08339	Aau08339 S. pneumo
743	32	52.5	474	7	ADD02817	Add02817 Petunia h	816	32	52.5	883	3	AAU90514	Aau90514 Streptoco
744	32	52.5	474	8	ADF38738	Adf38738 Petunia x	817	32	52.5	883	4	AAU38091	Aau38091 Streptoco
745	32	52.5	476	6	ABW20111	Abw20111 Novel hum	818	32	52.5	883	4	AAU37789	Aau37789 Streptoco
746	32	52.5	482	6	ABMG6852	Abmg6852 Photorhab	819	32	52.5	883	5	AAU97883	Aau97883 Streptoco
747	32	52.5	484	3	AAW82389	Aaw82389 C. tracho	820	32	52.5	883	6	ABU00937	Abu00937 S. pneumo
748	32	52.5	484	3	AAW42297	Aaw42297 Human ORF	821	32	52.5	883	6	ABU45915	Abu45915 Protein e
749	32	52.5	496	7	ADG42187	Adg42187 Human bra	822	32	52.5	883	8	ADH96823	Adh96823 S. pneumo
750	32	52.5	506	4	AAW25593	Aaw25593 Human pro	823	32	52.5	883	8	ADH96825	Adh96825 S. pneumo
751	32	52.5	514	3	AAW82391	Aaw82391 C. tracho	824	32	52.5	883	8	ADK46385	Adk46385 Streptoco
752	32	52.5	514	8	ADNI19937	Adni19937 Bacterial	825	32	52.5	883	9	AEC13178	Aec13178 Streptoco
753	32	52.5	534	8	ABM80561	Abm80561 Tumour-as	826	32	52.5	883	9	AEC13180	Aec13180 Streptoco

827	32	52.5	884	8	ADU69553	Adu69553 S agalact	900	31.5	51.6	2844	7	ADL35963	Adl35963 Human NOV
828	32	52.5	884	8	ADV88336	Adv88336 Streptoco	901	31.5	51.6	2845	5	ABG94631	Abg94631 Human NOV
829	32	52.5	884	8	ADV81756	Adv81756 Streptoco	902	31.5	51.6	2845	7	ADL35967	Adl35967 Human NOV
830	32	52.5	884	8	ADV79589	Adv79589 Streptoco	903	31.5	51.6	2877	5	ABG94630	Abg94630 Human NOV
831	32	52.5	888	8	ADR94329	Adr94329 Novel S.	904	31.5	51.6	2877	7	ADL35965	Adl35965 Human NOV
832	32	52.5	888	7	AEA58199	Aea58199 Streptoco	905	31.5	51.6	2995	7	ADL35961	Adl35961 Human NOV
833	32	52.5	912	7	ADD14183	Add14183 Human src	906	31.5	51.6	3415	9	AEE04798	Aee04798 Cancer-as
834	32	52.5	926	5	ADP69492	Adp69492 Human pol	907	31	50.8	9	3	AAV69597	AAv69597 Monoclonal
835	32	52.5	926	5	ADR40159	Adr40159 Human pro	908	31	50.8	9	3	AAV50888	AAv50888 Antibody
836	32	52.5	926	8	ADP54550	Adp54550 Human PRO	909	31	50.8	9	9	ADU70618	Adu70618 Human hep
837	32	52.5	926	9	ADX05516	Adx05516 Cyclin-de	910	31	50.8	10	9	ADV59607	Adv59607 G protein
838	32	52.5	926	9	ADY15498	Ady15498 PRO polyp	911	31	50.8	12	2	AAW38089	AAw38089 PEPpy mot
839	32	52.5	931	6	ABU31995	Abu31995 Protein e	912	31	50.8	12	9	ADB49350	Adb49350 Biocinylin
840	32	52.5	951	4	AAU34872	Aau34872 E. coli c	913	31	50.8	14	9	ADV59626	Adv59626 G protein
841	32	52.5	951	4	AAU38260	Aau38260 Salmonell	914	31	50.8	15	9	ADU71085	Adu71085 Human hep
842	32	52.5	951	6	ABU15050	Abu15050 Protein e	915	31	50.8	16	5	ABB79701	Abb79701 Integrin
843	32	52.5	951	6	ABU48273	Abu48273 Protein e	916	31	50.8	16	9	ADV59623	Adv59623 G protein
844	32	52.5	951	6	ABU28309	Abu28309 Protein e	917	31	50.8	16	9	ADZ68070	Adz68070 MAP kinase
845	32	52.5	953	6	ABU49570	Abu49570 Protein e	918	31	50.8	17	3	AAW25992	AAw25992 Human Igg
846	32	52.5	962	6	ABU41155	Abu41155 Protein e	919	31	50.8	17	4	AAU16717	Aau16717 Peptide E
847	32	52.5	965	6	ABU49983	Abu49983 Protein e	920	31	50.8	17	5	ABJ00247	Abj00247 Human Igg
848	32	52.5	967	7	ABO62249	Abu62249 Klebsiell	921	31	50.8	23	7	ADB87673	Adb87673 Human neu
849	32	52.5	975	7	ADF05470	Adf05470 Bacterial	922	31	50.8	27	2	AAW14660	AAw14660 Integrin
850	32	52.5	979	6	ABM68329	Abm68329 Photorhab	923	31	50.8	27	5	AAW50485	AAw50485 Integrin
851	32	52.5	982	8	ADS29846	Ads29846 Bacterial	924	31	50.8	27	5	AAW50503	AAw50503 Integrin
852	32	52.5	989	7	ADM29380	Adm29380 Human nov	925	31	50.8	27	5	AAW50504	AAw50504 Integrin
853	32	52.5	1006	6	ABP80935	Abp80935 N. gonorr	926	31	50.8	37	7	ADC88007	Adc88007 Ribosomal
854	32	52.5	1010	9	ABN91379	Abn91379 Microbial	927	31	50.8	38	5	ABP28976	Abp28976 Streptoco
855	32	52.5	1019	8	ADN22389	Adn22389 Bacterial	928	31	50.8	40	4	AAW66762	AAw66762 Beta2 sub
856	32	52.5	1047	5	ABG61533	Abg61533 Human tra	929	31	50.8	40	5	ABG60572	Abg60572 Selective
857	32	52.5	1047	7	ADG63021	Adg63021 Human pro	930	31	50.8	40	5	ABB79719	Abb79719 Integrin
858	32	52.5	1047	7	ADE63017	Ade63017 Human Pro	931	31	50.8	42	3	AAV90537	AAv90537 Conus sul
859	32	52.5	1047	8	ADQ88701	Adq88701 Human ATP	932	31	50.8	42	3	AAV90534	AAv90534 Conus atr
860	32	52.5	1082	9	AEE02346	Aee02346 Human her	933	31	50.8	42	3	AAV32353	AAv32353 Human Clq
861	32	52.5	1093	2	AAW41001	Aaw41001 Human myo	934	31	50.8	46	2	AAW80964	AAw80964 Integrin
862	32	52.5	1132	4	ABG07312	Abg07312 Novel hum	935	31	50.8	46	2	AAW65886	AAw65886 Cytoplasm
863	32	52.5	1272	6	ABP72190	Abp72190 Plasmodiu	936	31	50.8	46	2	AAW48690	AAw48690 Amino aci
864	32	52.5	1429	5	ABP35654	Abp35654 Fungal ZB	937	31	50.8	46	2	AAW43097	AAw43097 Human int
865	32	52.5	1433	5	ABP35624	Abp35624 Fungal ZB	938	31	50.8	46	3	AAW26128	AAw26128 Beta2-int
866	32	52.5	1610	7	ADB46133	Adb46133 Plasmodiu	939	31	50.8	46	3	ABO54611	Abu54611 Human gen
867	32	52.5	1804	6	ABR53154	AbR53154 Protein s	940	31	50.8	47	3	AAW26129	AAw26129 Beta2-int
868	32	52.5	1804	7	ADK63138	Adk63138 Disease t	941	31	50.8	54	4	AAU62166	AAu62166 Propionib
869	32	52.5	2000	4	ABW61853	Abw61853 Myxoma vi	942	31	50.8	54	6	ABW58685	Abw58685 Propionib
870	32	52.5	2197	8	ADK16573	Adk16573 Nanoarcha	943	31	50.8	57	5	ABB81140	Abb81140 Human ret
871	32	52.5	4655	2	AAW43312	Aaw43312 Human pla	944	31	50.8	63	6	AAU42527	AAu42527 Propionib
872	31.5	51.6	85	2	AAW55353	Aaw55353 H. pylori	945	31	50.8	63	6	ABM39046	Abm39046 Propionib
873	31.5	51.6	244	4	AAW98655	Aaw98655 Caspase-7	946	31	50.8	71	5	AAV79388	AAv79388 EGF-like
874	31.5	51.6	303	2	AAW15262	Aaw15262 Apoptotic	947	31	50.8	71	5	AAE17041	AAe17041 Human G p
875	31.5	51.6	303	2	AAW15247	Aaw15247 Cysteine	948	31	50.8	71	8	ADL14217	Adl14217 Novel hum
876	31.5	51.6	303	2	AAV21721	AAv21721 Amino aci	949	31	50.8	77	2	AAW88749	AAw88749 S. aureus
877	31.5	51.6	303	4	AAE00604	Aae00604 Human cas	950	31	50.8	77	4	AAU63452	AAu63452 Propionib
878	31.5	51.6	303	5	ABJ01222	Abj01222 Human cas	951	31	50.8	77	5	ABP02587	Abp02587 Human ORF
879	31.5	51.6	303	5	ABO9299	Abu9299 Human cas	952	31	50.8	77	6	ABM59971	Abm59971 Propionib
880	31.5	51.6	303	5	ABO9297	Abu9297 Human cas	953	31	50.8	82	7	ADF30910	Adf30910 Soil meta
881	31.5	51.6	303	6	AAO19868	Aao19868 Bacteriop	954	31	50.8	84	3	AAW61260	AAw61260 Arabidops
882	31.5	51.6	303	8	ADOL6852	Adol6852 Human cas	955	31	50.8	84	3	AAW57207	AAw57207 Arabidops
883	31.5	51.6	305	9	ADV90184	Adv90184 Human pro	956	31	50.8	85	2	AAW04170	AAw04170 Flea calr
884	31.5	51.6	305	9	ADV69228	Adv69228 Human Cys	957	31	50.8	88	4	ABG25271	Abg25271 Novel hum
885	31.5	51.6	327	10	AEP11566	Aep11566 Tomato me	958	31	50.8	89	5	ADK35323	Adk35323 Novel hum
886	31.5	51.6	330	4	ABB69802	Abb69802 Drosophil	959	31	50.8	91	3	AAW57206	AAw57206 Arabidops
887	31.5	51.6	330	8	ADS96506	Ads96506 Drosophil	960	31	50.8	92	4	AAW51745	AAw51745 Human FSH
888	31.5	51.6	336	2	AAW09300	AAw09300 Human cas	961	31	50.8	96	3	AAW57205	AAw57205 Arabidops
889	31.5	51.6	341	5	AAW95830	Aaw95830 Human int	962	31	50.8	96	3	AAW61259	AAw61259 Arabidops
890	31.5	51.6	1118	4	AAW50209	Aaw50209 Human fib	963	31	50.8	99	8	ADR96565	Adr96565 Novel S.
891	31.5	51.6	1600	5	ABG66678	Abg66678 Human nov	964	31	50.8	99	8	AAW60435	AAw60435 Streptoco
892	31.5	51.6	1716	9	AEE04794	Aee04794 Cancer-as	965	31	50.8	101	2	AAW90939	AAw90939 D3D4 poly
893	31.5	51.6	1910	7	AED04796	Aed04796 Cancer-as	966	31	50.8	102	3	AAW57676	AAw57676 Arabidops
894	31.5	51.6	2676	9	ADL35971	Adl35971 Human NOV	967	31	50.8	102	3	AAW10138	AAw10138 Arabidops
895	31.5	51.6	2695	5	ABG94632	Abg94632 Human NOV	968	31	50.8	103	4	AAW35877	AAw35877 C. pneumo
896	31.5	51.6	2695	7	ADL35969	Adl35969 Human NOV	969	31	50.8	109	4	ABG10963	Abg10963 Novel hum
897	31.5	51.6	2757	5	ABG94633	Abg94633 Human NOV	970	31	50.8	113	2	AAW90937	AAw90937 D3D4 poly
898	31.5	51.6	2809	5	AAW66169	Aaw66169 Human fib	971	31	50.8	113	4	ABB62220	Abb62220 Drosophil
899	31.5	51.6	2844	5	ABG94629	Abg94629 Human NOV	972	31	50.8	115	5	ADK36270	Adk36270 Novel hum

973 31 50.8 116 5 ABP06792
974 31 50.8 117 5 ADK34293
975 31 50.8 123 8 AAX95695
976 31 50.8 124 8 AAR90940
977 31 50.8 126 8 ADS24055
978 31 50.8 127 8 ADY12815
979 31 50.8 128 2 AAR90941
980 31 50.8 129 3 AAG10137
981 31 50.8 130 2 AAR58806
982 31 50.8 130 2 AAR58807
983 31 50.8 130 2 AAR91202
984 31 50.8 130 2 AAR91201
985 31 50.8 131 2 AAR95447
986 31 50.8 131 8 ADX95773
987 31 50.8 132 4 AAU29691
988 31 50.8 133 4 AAU50531
989 31 50.8 133 6 ABM47050
990 31 50.8 134 2 AAY29089
991 31 50.8 134 3 AAY29088
992 31 50.8 135 3 AAG57675
993 31 50.8 137 2 AAY44012
994 31 50.8 140 3 AAG57674
995 31 50.8 141 4 AAU08682
996 31 50.8 141 7 ADB32031
997 31 50.8 145 4 AAU30715
998 31 50.8 147 8 ADO26501
999 31 50.8 153 3 AAG10136
1000 31 50.8 154 8 ADY11312

ALIGNMENTS

RESULT 1
ADR88212
ID ADR88212 standard; peptide; 10 AA.
XX AC ADR88212;
XX DT 18-NOV-2004 (first entry)
XX DE Human heparanase epitope pep38.
XX KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
KW human; heparanase; enzyme; epitope.
XX OS Homo sapiens.
XX PN US2004170631-A1.
XX PD 02-SEP-2004.
XX PF 28-NOV-2003; 2003US-00722502.
XX PR 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
PR 22-AUG-2003; 2003US-00645659.
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VGEL/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX WPI; 2004-625084/60.
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX Claim 7; SEQ ID NO 6; 58pp; English.
XX The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an antibody or its portion
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease), for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is human heparanase epitope.
XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 8; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
| | | | | | | | | |
Db 1 CTNTDNPYK 10

RESULT 2
ADT78179
ID ADT78179 standard; peptide; 10 AA.
XX AC ADT78179;
XX DT 13-JAN-2005 (first entry)
XX DE Functional peptide epitope of human heparanase, pep38.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 XX vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.

OS Homo sapiens.

XX US2004213789-A1.

XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

XX 01-MAY-1998; 98US-00071739.

XX 04-NOV-1998; 98US-00186200.

XX 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 39; SEQ ID NO 6; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumor cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents a functional peptide
 CC epitope of human heparanase.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 8; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

DB 1 CTNTDNPYK 10

RESULT 3

AEA42428

ID AEA42428 standard; peptide; 10 AA.

XX AEA42428;

XX 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:6.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotrophic;
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

OS Homo sapiens.

XX AU2004201462-A1.

XX 06-MAY-2004.

XX 08-APR-2004; 2004AU-00201462.

XX 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.

XX Claim 7; SEQ ID NO 6; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (I) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents a human
 CC heparanase epitope peptide, which is used in the exemplification of the
 CC present invention.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 9; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

```

Db      1 CTNTDNPYK 10
|||||
RESULT 4
ADU71246
ID ADU71246 standard; peptide; 15 AA.
XX
AC ADU71246;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human heparanase peptide SEQ ID NO:931.
XX
KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;
KW immunostimulant; cytostatic; immune disorder; metastasis.
XX
OS Homo sapiens.
XX
PN EP1479764-A1.
XX
PD 24-NOV-2004.
XX
PF 19-MAY-2003; 2003EP-00011038.
XX
PR 19-MAY-2003; 2003EP-00011038.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.
XX
PI Schirmacher V, Beckhove P, Sommerfeldt N;
XX
WPI; 2005-014847/02.
XX
PT New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
PT molecule or its functional derivative, useful for preparing a medicament
PT for inducing an immune response or for treating metastatic tumors.
XX
PS Disclosure; SEQ ID NO 931; 269pp; English.
XX
CC The invention relates to a novel heparanase peptide that binds to a human
CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
CC its functional derivative. A peptide of the invention has immunostimulant
CC and cytostatic activity, and is used in a vaccine. The heparinase peptide
CC is useful for preparing a medicament which induces an immune response or
CC for treating metastatic tumors. The present sequence represents a
CC heparinase peptide of the invention.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 61; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
|||||
Db      2 CTNTDNPYK 15
|||||
RESULT 6
ADU71087
ID ADU71087 standard; peptide; 15 AA.
XX
AC ADU71087;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human heparanase peptide SEQ ID NO:772.
XX
KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;
KW immunostimulant; cytostatic; immune disorder; metastasis.
XX
OS Homo sapiens.
XX
PN EP1479764-A1.
XX
PD 24-NOV-2004.
XX
PF 19-MAY-2003; 2003EP-00011038.
XX
PR 19-MAY-2003; 2003EP-00011038.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.
XX
PI Schirmacher V, Beckhove P, Sommerfeldt N;
XX
WPI; 2005-014847/02.
XX
PT New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
PT molecule or its functional derivative, useful for preparing a medicament
PT for inducing an immune response or for treating metastatic tumors.
XX
SQ Sequence 15 AA;
Query Match 100.0%; Score 61; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
|||||
Db      2 CTNTDNPYK 11
|||||
RESULT 5
ADU70983
ID ADU70983 standard; peptide; 15 AA.
XX
AC ADU70983;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human heparanase peptide SEQ ID NO:668.
XX
KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;
KW immunostimulant; cytostatic; immune disorder; metastasis.
XX
OS Homo sapiens.

```

PT for inducing an immune response or for treating metastatic tumors.

PS Disclosure; SEQ ID NO 772; 269pp; English.

XX The invention relates to a novel heparanase peptide that binds to a human
 CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
 CC its functional derivative. A peptide of the invention has immunostimulant
 CC and cytostatic activity, and is used in a vaccine. The heparinase peptide
 CC is useful for preparing a medicament which induces an immune response or
 CC for treating metastatic tumors. The present sequence represents a
 CC heparinase peptide of the invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 61; DB 9; Length 15;

Best Local Similarity 100.0%; Pred. No. 0.0024;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

Db 3 CTNTDNPYK 12

RESULT 7

ADU71245

ID ADU71245 standard; peptide; 15 AA.

XX AC ADU71245;

XX 10-FEB-2005 (first entry)

XX Human heparanase peptide SEQ ID NO:930.

XX enzymes; heparinase; vaccine; human leukocyte antigen; HLA;
 KW immunostimulant; cytostatic; immune disorder; metastasis.
 KW Homo sapiens.

OS EP1479764-A1.

XX 24-NOV-2004.

XX 19-MAY-2003; 2003EP-00011038.

XX 19-MAY-2003; 2003EP-00011038.

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.

XX Schirmacher V, Beckhove P, Sommerfeldt N;

XX WPI; 2005-014847/02.

XX New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
 PT molecule or its functional derivative, useful for preparing a medicament
 PT for inducing an immune response or for treating metastatic tumors.
 XX Disclosure; SEQ ID NO 930; 269pp; English.

XX The invention relates to a novel heparanase peptide that binds to a human
 CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
 CC its functional derivative. A peptide of the invention has immunostimulant
 CC and cytostatic activity, and is used in a vaccine. The heparinase peptide
 CC is useful for preparing a medicament which induces an immune response or
 CC for treating metastatic tumors. The present sequence represents a
 CC heparinase peptide of the invention.

XX Sequence 15 AA;

Query Match

Best Local Similarity 100.0%; Score 61; DB 9; Length 15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

Db 4 CTNTDNPYK 13

RESULT 8

ADR88207

ID ADR88207 standard; protein; 386 AA.

XX ADR88207;

XX 18-NOV-2004 (first entry)

XX Human mature heparanase 45 kDa major subunit.

XX Targeted drug delivery; inflammatory disorder; wound; scar;
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;
 KW non-solid cancer; haematopoietic malignancy; lymphocytic leukaemia;
 KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;
 KW haemangiosarcoma; Kaposi's sarcoma; human; heparanase; enzyme.

OS Homo sapiens.

XX US2004170631-A1.

XX 02-SEP-2004.

XX 28-NOV-2003; 2003US-00722502.

XX 02-SEP-1997; 97US-00922170.

XX 01-MAY-1998; 98US-00071739.

XX 04-NOV-1998; 98US-00186200.

XX 19-FEB-2003; 2003US-00368044.

XX 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

XX (PERE/) PERETZ T.

XX (MIRO/) MIRON D.

XX (SHLO/) SHLOMI Y.

XX (PECK/) PECKER I.

XX (AVAL/) AVAL-HERSHKOVITZ M.

XX (FEIN/) FEINSTEIN E.

XX (VDEL/) VAN GELDER J M.

XX (VLOD/) VLODAVSKY I.

XX (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

XX Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

XX Friedmann Y;

XX WPI; 2004-625084/60.

XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
 PT useful for treating heparanase-associated conditions such as inflammation
 PT or cancer, comprises administering a drug and an anti-heparanase antibody
 PT complex.

PS Claim 2; SEQ ID NO 1; 59pp; English.

XX The invention relates to a method of targeted drug delivery to a tissue
 CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an antibody or its portion
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,

CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature
 CC heparanase.
 XX
 SQ Sequence 386 AA;
 Query Match 100.0%; Score 61; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 0.06;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 DB 280 CTNTDNPYK 289
 RESULT 9
 ADT78174
 ID ADT78174 standard; protein; 386 AA.
 XX AC
 XX ADT78174;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE 45kDa subunit of mature processed human heparanase dimer.
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnery; arteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
 XX
 OS Homo sapiens.
 XX
 XX US2004213789-A1.
 XX
 XX 28-OCT-2004.
 XX
 XX 22-AUG-2003; 2003US-00645659.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX
 XX (YACO/) YACOBY-ZEEVI O.
 XX (PERE/) PERETZ T.
 XX (MIRO/) MIRON D.
 XX (SHLO/) SHLOMI Y.
 XX (PECK/) PECKER I.
 XX (AYAL/) AYAL-HERSHKOVITZ M.
 XX (FEIN/) FEINSTEIN E.
 XX (GELD/) GELDER J M V.
 XX (VL0D/) VL0DAVSKY I.
 XX (FRIE/) FRIEDMANN Y.
 XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 XX Friedmann Y;
 DR WPI; 2004-774790/76.
 XX
 PT New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX
 PS Claim 5; SEQ ID NO 1; 68pp; English.
 XX
 CC The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of
 CC mature processed human heparanase dimer.
 XX
 SQ Sequence 386 AA;
 Query Match 100.0%; Score 61; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 0.06;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 DB 280 CTNTDNPYK 289
 RESULT 10
 ADY27057
 ID ADY27057 standard; protein; 386 AA.
 XX
 XX AC
 XX ADY27057;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 DE Heparanase inhibitor protein #1.
 XX
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW heparanase modulator; enzyme purification.
 XX
 XX OS Homo sapiens.
 XX
 XX WO2005016227-A2.
 XX
 XX 24-FEB-2005.
 PD
 XX 12-AUG-2004; 2004WO-IL000744.
 PF
 XX 14-AUG-2003; 2003US-0494800P.
 XX
 PR 12-JAN-2004; 2004US-0535492P.
 PR

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PI Van-Gelder JM, Miron D;
XX DR WPI; 2005-182203/19.
XX PT Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.
XX PS Claim 55; SEQ ID NO 33; 211pp; English.
XX CC The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the
XX invention.
XX SQ Sequence 386 AA;

Query Match 100.0%; Score 61; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
Db 280 CTNTDNPYK 289
|||||

RESULT 11
ADZ18995
ID ADZ18995 standard; protein; 386 AA.
XX AC ADZ18995;
XX DT 16-JUN-2005 (first entry)
XX DE Human heparanase consensus cleavage site #2.
XX KW Enzyme engineering; heparanase; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX OS Homo sapiens.
XX PN WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
XX PR 20-JAN-2004; 2004US-0537729P.
XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

XX DR WPI; 2005-273382/28.
XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX PS Disclosure; SEQ ID NO 16; 65pp; English.
XX CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a human heparanase
CC consensus cleavage site used in the scope of the invention.
XX SQ Sequence 386 AA;

Query Match 100.0%; Score 61; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
Db 280 CTNTDNPYK 289
|||||

RESULT 12
AEA42423
ID AEA42423 standard; protein; 386 AA.
XX AC AEA42423;
XX DT 28-JUL-2005 (first entry)
XX DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.
XX KW antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
KW angiogenesis disorder; cancer; tumor; metastasis.
XX OS Homo sapiens.
XX PN AU2004201462-A1.
XX PD 06-MAY-2004.
XX PF 08-APR-2004; 2004AU-00201462.
XX PR 08-APR-2004; 2004AU-00201462.
XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinstein E;
XX DR WPI; 2005-173343/19.
XX PT Novel isolated antibody capable of specifically binding to epitope of
PT heparanase protein, useful for preventing and treating heparanase-related
PT disorder such as inflammatory disorder, scars, autoimmune conditions or
PT angiogenesis.
XX PS Claim 2; SEQ ID NO 1; 260pp; English.
XX CC The invention relates to an isolated antibody or its portion (I) capable
CC of specifically binding to an epitope of a heparanase protein. Also

described: (1) a cell line (II) for producing a monoclonal antibody or its portion, comprising a cell line for producing (I); (2) a pharmaceutical composition comprising (I) and a carrier; and (3) an affinity medium (III) for binding human heparanase polypeptides, comprising (I) immobilized to a chemically inert, insoluble carrier. (I) is useful for treating a subject suffering from a pathological condition, which involves administering (I) to the subject. (I) is useful for preventing and treating heparanase-related disorder or condition chosen from inflammatory disorder, wound, scar, vasculopathy, autoimmune condition, angiogenesis, cell proliferation, cancerous condition, tumor cell proliferation, invasion of circulating tumor cells and metastatic disease. (I) is useful for detecting the presence of heparanase polypeptide in a sample. (I) is useful for detecting heparanase-related disease or condition in a subject such as vertebrate, preferably mammal e.g., human. The heparanase-related disorder or condition further includes renal disease or disorder chosen from diabetic nephropathy, glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome and renal cell carcinoma. The present sequence represents the 45 kDa subunit of the human mature processed heparanase dimer, which is used in the exemplification of the present invention.

Query Match 100.0%; Score 61; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
| | | | | | | |
DB 280 CTNTDNPYK 289

RESULT 13

ADV27061
ID ADV27061 standard; protein; 460 AA.

AC ADV27061;
XX 05-MAY-2005 (first entry)
DE Heparanase inhibitor protein #4.
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.

XX Homo sapiens.
OS
XX WO2005016227-A2.
PN
XX 24-FEB-2005.
PD
XX 12-AUG-2004; 2004WO-IL000744.
PF
XX 14-AUG-2003; 2003US-0494800P.
PR
XX 12-JAN-2004; 2004US-0535492P.
PR
XX (INSI-) INSTIGHT BIOPHARMACEUTICALS LTD.

PA Van-Gelder JM, Miron D;
XX WPI; 2005-182203/19.
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.

XX Disclosure; SEQ ID NO 37; 211pp; English.
PS
XX The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on

CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the
CC invention.

XX Sequence 460 AA;

Query Match 100.0%; Score 61; DB 9; Length 460;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
| | | | | | | |
DB 354 CTNTDNPYK 363

RESULT 14

AE8B7589
ID AE8B7589 standard; protein; 486 AA.

XX AE8B7589;
AC
XX 06-OCT-2005 (first entry)
DT Human heparanase 65delta20 deletion mutant.
DE
XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; muten.

XX Homo sapiens.
OS
XX Synthetic.

XX WO2005071070-A2.

XX 04-AUG-2005.

XX 20-JAN-2005; 2005WO-IL0000068.

XX 22-JAN-2004; 2004IL-00160025.

XX 28-JUL-2004; 2004US-00901943.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Ilan N, Levy-Adam F;

XX WPI; 2005-564219/57.

XX N-PSDB; AE8B7588.

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.

XX Claim 107; SEQ ID NO 31; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the

CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX Sequence 486 AA;
 SQ

Query Match 100.0%; Score 61; DB 9; Length 486;
 Best Local Similarity 100.0%; Pred. No. 0.075;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 DB 380 CTNTDNPYK 389

RESULT 15
 ADZ18996
 ID ADZ18996 standard; protein; 492 AA.
 XX
 AC ADZ18996;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE Hep106 construct protein.
 XX
 KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 XX
 PI Lahm A, Nardella C, Pallaooro M, Steinkuhler C;
 XX
 DR WPI; 2005-273382/28.
 DR N-PSDB; ADZ18997.
 XX
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 PS Example 2; SEQ ID NO 17; 65pp; English.
 XX
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or

CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hep106 construct protein
 CC used in the scope of the invention.
 XX Sequence 492 AA;
 SQ

Query Match 100.0%; Score 61; DB 9; Length 492;
 Best Local Similarity 100.0%; Pred. No. 0.076; 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 DB 386 CTNTDNPYK 395

RESULT 16
 AEB87562
 ID AEB87562 standard; protein; 493 AA.
 XX
 AC AEB87562;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE Human heparanase 65delta15 deletion mutant.
 XX
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO2005071070-A2.
 XX
 PD 04-AUG-2005.
 XX
 XX 20-JAN-2005; 2005WO-IL000068.
 PF
 XX 22-JAN-2004; 2004IL-00160025.
 PR
 PR 28-JUL-2004; 2004US-00901943.
 XX
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
 XX
 PI Vlodavsky I, Ilan N, Levy-Adam F;
 XX
 DR WPI; 2005-564219/57.
 DR N-PSDB; AEB87561.
 XX
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.
 XX
 PS Claim 105; SEQ ID NO 4; 167pp; English.
 XX
 CC The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,

CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
 XX
 SQ Sequence 493 AA;

Query Match 100.0%; Score 61; DB 9; Length 493;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||

Db 387 CTNTDNPYK 396

RESULT 17

ADZ18999
 ID ADZ18999 standard; protein; 495 AA.

XX AC ADZ18999;

DT 16-JUN-2005 (first entry)

DE Hep109 construct protein.

XX Enzyme engineering; heparanase; hepl09; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.

OS Synthetic.

XX WO2005030962-A1.

PN 07-APR-2005.

XX 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

PR 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

XX WPI; 2005-273382/28.

DR N-PSDB; ADZ18998.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.

XX Example 2; SEQ ID NO 20; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepl09 construct protein
 CC used in the scope of the invention.

XX Sequence 495 AA;

Query Match 100.0%; Score 61; DB 9; Length 495;
 Best Local Similarity 100.0%; Pred. No. 0.077;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||

Db 389 CTNTDNPYK 398

RESULT 18

AEB87587

ID AEB87587 standard; protein; 497 AA.

XX AC AEB87587;

XX 06-OCT-2005 (first entry)

XX Human heparanase 65delta10 deletion mutant.

XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
 KW leukemia; lymphoma; dermatological disease; hematological disease;
 KW immune disorder; inflammation; antiinflammatory; renal disease;
 KW nephrotropic; endocrine disease; genitourinary disease;
 KW autoimmune disease; immunosuppressive; drug screening; mutein.

OS Homo sapiens.

OS Synthetic.

XX WO2005071070-A2.

XX 04-AUG-2005.

XX 20-JAN-2005; 2005WO-IL000068.

XX 22-JAN-2004; 2004IL-00160025.

PR 28-JUL-2004; 2004US-00901943.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodayvsky I, Ilan N, Levy-Adam F;

XX WPI; 2005-564219/57.

DR N-PSDB; AEB87586.

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney
 PT disorder or autoimmune disorder.

XX Claim 106; SEQ ID NO 29; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,
 CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the
 CC native protein. The recombinant protein is deficient of heparanase
 CC endoglycosidase catalytic activity. The invention relates to amino acid
 CC sequences derived from the N-terminus region of the 50 kDa subunit of
 CC heparanase, particularly in the regions between amino acid residues 158-
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
 CC heparin-binding domains. The invention also provides an antibody directed
 CC to these sequences, in particular the 158-171 peptide, and compositions
 CC and uses of this antibody as a heparanase inhibitor. A screening method
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
 CC compositions comprising (i) a peptide derived from the N-terminus region
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
 CC peptide, or (iii) an antibody which specifically recognizes the peptide
 CC are used for the inhibition of heparanase catalytic activity associated
 CC with an inflammatory disorder, kidney disease, autoimmune disease,
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
 CC with a malignant proliferative disorder, especially a solid or non-solid
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.

XX Sequence 497 AA;

Query Match 100.0%; Score 61; DB 9; Length 497;
 Best Local Similarity 100.0%; Pred. No. 0.077;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

|||||

Db 391 CTNTDNPYK 400

RESULT 19
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX
XX ADZ19000;
AC ADZ19000;
DT 16-JUN-2005 (first entry)
XX
DE HepG3 construct protein.
XX
KW Enzyme engineering; heparanase; hepgS3; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
XX WO2005030962-A1.
PN
XX
PD 07-APR-2005.
XX
DT 17-SEP-2004; 2004WO-EP010517.
XX
DE HepG3 construct protein.
XX
KW Enzyme engineering; heparanase; hepgS3; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
XX WO2005030962-A1.
PN
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
PA
XX
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
PI
XX
XX WPI; 2005-273382/28.
DR
XX
XX N-PSDB; ADZ19001.
DR
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 21; 65pp; English.
PS
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepgS3 construct protein
CC used in the scope of the invention.
XX
XX Sequence 501 AA;
SQ
Query Match 100.0%; Score 61; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNTDNPYK 10
Db 395 CTNTDNPYK 404
RESULT 20
ADZ19005
ID ADZ19005 standard; protein; 507 AA.
XX
XX ADZ19005;
AC
DT 16-JUN-2005 (first entry)
XX
DE HepG6 construct protein.
XX
KW Enzyme engineering; heparanase; hepgS6; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.

XX
OS Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
XX
XX N-PSDB; ADZ19003.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 26; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepgS6 construct protein
CC used in the scope of the invention.
XX
XX Sequence 507 AA;
SQ
Query Match 100.0%; Score 61; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNTDNPYK 10
Db 401 CTNTDNPYK 410
RESULT 21
ADZ27058
ID ADZ27058 standard; protein; 508 AA.
XX
XX ADY27058;
AC
XX
XX 05-MAY-2005 (first entry)
DT
XX
XX Human inactive heparanase protein.
DE
XX
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW protease; enzyme; enzyme purification.
XX
XX Homo sapiens.
OS
XX
XX WO2005016227-A2.
PN
XX
XX 24-FEB-2005.
PD
XX
XX 12-AUG-2004; 2004WO-IL000744.
XX
XX 14-AUG-2003; 2003US-0494800P.
PR
XX 12-JAN-2004; 2004US-0535492P.
PR
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA

XX PI Van-Gelder JM, Miron D;
 XX WPI; 2005-182203/19.
 XX
 XX Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.
 XX
 XX Claim 257; SEQ ID NO 34; 211pp; English.
 XX
 XX The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical
 CC carrier or diluent, a method of identifying a protease activator of
 CC heparanase, a protease substrate mimetic comprising a peptide
 CC representing a subset or all substrate residues or cleavage sites of
 CC human heparanase or an equivalent non-human heparanase, a method of
 CC producing active heparanase and a method of modulating an adhesion
 CC activity of heparanase. The composition and methods are useful for
 CC modulating heparanase activation and for treating heparanase-associated
 CC diseases or disorders such as cancer, inflammation, cardiovascular
 CC diseases, neurological diseases or viral infections. This sequence
 CC represents a human inactive heparanase protein used in the scope of the
 CC invention.
 XX
 XX Sequence 508 AA;
 SQ
 Query Match 100.0%; Score 61; DB 9; Length 508;
 Best Local Similarity 100.0%; Pred. No. 0.079;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 Db |||||
 402 CTNTDNPYK 411
 RESULT 22
 ADZ19006
 ID ADZ19006 standard; protein; 526 AA.
 XX
 XX ADZ19006;
 AC
 DT 16-JUN-2005 (first entry)
 DE Hephyluro construct protein.
 XX
 XX Enzyme engineering; heparanase; hephyluro; metastasis;
 KW autoimmune disease; inflammation; neoplasm; immune disorder;
 KW antiinflammatory; cytostatic; immunosuppressive; enzyme.
 XX
 XX Synthetic.
 OS
 XX WO2005030962-A1.
 PN
 XX 07-APR-2005.
 PD
 XX 17-SEP-2004; 2004WO-EP010517.
 PF
 XX 26-SEP-2003; 2003US-0506479P.
 PR
 XX 20-JAN-2004; 2004US-0537729P.
 XX
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
 PA Lahm A, Nardella C, Pallao M, Steinkuhler C;
 XX WPI; 2005-273382/28.
 XX
 XX

DR N-PSDB; ADZ19007.
 XX
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 XX Example 2; SEQ ID NO 27; 65pp; English.
 XX
 XX The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hephyluro construct
 CC protein used in the scope of the invention.
 XX
 XX Sequence 526 AA;
 SQ
 Query Match 100.0%; Score 61; DB 9; Length 526;
 Best Local Similarity 100.0%; Pred. No. 0.081;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 Db |||||
 420 CTNTDNPYK 429
 RESULT 23
 ABB07815
 ID ABB07815 standard; protein; 527 AA.
 XX
 XX ABB07815;
 AC
 XX 03-JUL-2002 (first entry)
 DT
 XX Chicken signal peptide/human heparanase chimeric protein sequence.
 DE
 XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.
 XX
 XX Synthetic.
 OS
 XX Gallus gallus.
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Peptide i. .19
 FT /note= "chicken heparanase signal peptide"
 FT Protein 20. .527
 FT /note= "human heparanase mature protein"
 XX
 XX US2002034810-A1.
 PN
 XX 21-MAR-2002.
 PD
 XX 16-AUG-2001; 2001US-00930218.
 PF
 XX 20-SEP-2000; 2000US-00666390.
 PR
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;
 PI
 XX WPI; 2002-338926/37.
 DR
 XX N-PSDB; ABL40753.
 XX
 XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.
 XX
 XX Disclosure; Page 26-28; 39pp; English.
 PS

XX The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a chicken signal peptide/human heparanase chimeric protein
 CC sequence
 XX SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 5; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
 Db 421 CTNTDNPYK 430
 |||||

RESULT 24
 ABW02018
 ID ABW02018 standard; protein; 527 AA.
 XX AC ABW02018;
 XX DT 12-FEB-2004 (first entry)
 XX DE Chimeric human-chicken heparanase protein.
 XX KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.
 XX OS Chimeric - Gallus gallus.
 OS Chimeric - Homo sapiens.
 XX PN US2003180788-A1.
 XX PD 25-SEP-2003.
 XX PF 08-MAY-2003; 2003US-00431438.
 XX PR 20-SEP-2000; 2000US-00666390.
 PR 16-AUG-2001; 2001US-00930218.
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
 XX WPI; 2003-843931/78.
 DR N-PSDB; AAD63532.
 XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.
 XX Example; Page 26-28; Opp; English.
 XX The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be

CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease). Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamine. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein
 XX SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 7; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
 Db 421 CTNTDNPYK 430
 |||||

RESULT 25
 ADO63825
 ID ADO63825 standard; protein; 527 AA.
 XX AC ADO63825;
 XX DT 26-AUG-2004 (first entry)
 XX DE Chimeric heparanase mutant E225A, SEQ ID:10.
 XX KW Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnerary; mutant; mutein.
 XX OS Homo sapiens.
 OS Gallus gallus.
 OS Synthetic.
 OS Chimeric.
 XX Key Location/Qualifiers
 FT Peptide 1..18 /note= "Chicken heparanase signal peptide"
 FT Region 19..527 /note= "Corresponds to residues 35-543 of human
 FT heparanase mutant E225A (SEQ ID NO:7)"
 FT Misc-difference 209 /note= "Ala replaces wild-type Glu (active site proton
 FT donor). Corresponds to residue 225 of human heparanase
 FT mutant E225A (SEQ ID NO:7)"
 FT Active-site 327 /note= "Active site nucleophile"
 FT WO2004048558-A2.
 XX PN 10-JUN-2004.
 XX PF 24-NOV-2003; 2003WO-IL000989.
 XX PR 24-NOV-2002; 2002IL-00153059.
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;
 XX WPI: 2004-450373/42.
 DR N-PSDB; ADO63819.
 XX New nucleic acid construct comprising heparanase-derived polypeptide,
 PT

PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 XX thrombasthenia, or Bernard-Soulier syndrome.
 PS Claim 10; SEQ ID NO 10; 128pp; English.
 XX
 CC The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase mutant E225A.
 XX
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 Db 421 CTNTDNPYK 430

RESULT 26
 ADO63827
 ID ADO63827 standard; protein; 527 AA.
 AC ADO63827;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Chimeric heparanase mutant E225A/E343A, SEQ ID:12.
 XX
 KW Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnerable; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Gallus gallus.
 OS Synthetic.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..18
 FT /notes= "Chicken heparanase signal peptide"
 FT Region 19..527
 FT /notes= "Corresponds to residues 35-543 of human
 FT heparanase mutant E225A/E343A (SEQ ID NO:9)"
 FT

Misc-difference 209
 /note= "Ala replaces wild-type Glu (active site proton
 donor). Corresponds to residue 225 of human heparanase
 mutant E225A/E343A (SEQ ID NO:9)"
 FT
 Misc-difference 327
 /note= "Ala replaces wild-type Glu (active site
 nucleophile). Corresponds to residue 343 of human
 heparanase mutant E225A/E343A (SEQ ID NO:9)"
 FT
 WO2004048558-A2.
 PD 10-JUN-2004.
 XX 24-NOV-2003; 2003WO-IL000989.
 XX 24-NOV-2002; 2002IL-00153059.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
 PI WPI; 2004-450373/42.
 DR N-PSDB; ADO63821.
 DR
 DR
 XX New nucleic acid construct comprising heparanase-derived polypeptide,
 FT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 FT thrombasthenia, or Bernard-Soulier syndrome.
 XX
 PS Claim 10; SEQ ID NO 12; 128pp; English.
 XX
 CC The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a chimeric
 CC protein comprising the signal peptide of chicken heparanase and residues
 CC 35-543 of the human heparanase double mutant E225A/E343A.
 XX
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 8; Length 527;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 Db 421 CTNTDNPYK 430

RESULT 27
 ADO63826
 ID ADO63826 standard; protein; 527 AA.
 XX

AC ADO63826;
XX 26-AUG-2004 (first entry)
XX Chimeric heparanase mutant E343A, SEQ ID:11.
XX Human; chicken; heparanase; heparanase-derived protein;
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein.
XX Homo sapiens.
OS Gallus gallus.
OS Synthetic.
OS Chimeric.
XX
XX Key Location/Qualifiers
FT Peptide 1..18
FT /note= "Chicken heparanase signal peptide"
FT Region 19..527
FT /note= "Corresponds to residues 35-543 of human
FT heparanase mutant E343A (SEQ ID NO:8)"
FT Active-site 209
FT /note= "Active site proton donor"
FT Misc-difference 327
FT /note= "Ala replaces wild-type Glu (active site
FT nucleophile)". Corresponds to residue 343 of human
FT heparanase mutant E343A (SEQ ID NO:8)"
XX
XX W02004048558-A2.
XX
XX 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-IL000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
PR
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
PI WPI; 2004-450373/42.
DR N-PSDB; ADO63820.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Claim 10; SEQ ID NO 11; 128pp; English.
PS
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be use in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet

CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a chimeric
CC protein comprising the signal peptide of chicken heparanase and residues
CC 35-543 of the human heparanase mutant E343A.
XX
SQ Sequence 527 AA;
Query Match 100.0%; Score 61; DB 8; Length 527;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNTDNPYK 10
DB 421 CTNTDNPYK 430
RESULT 28
ADZ19004
ID ADZ19004 standard; protein; 527 AA.
XX
AC ADZ19004;
XX
DT 16-JUN-2005 (first entry)
XX
DE HepGS4 construct protein.
XX
KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
XX WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
DR N-PSDB; ADZ19002.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 25; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepGS4 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 527 AA;
Query Match 100.0%; Score 61; DB 9; Length 527;
Best Local Similarity 100.0%; Pred. No. 0.082;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 Db 421 CTNTDNPYK 430

RESULT 29
 AAY34173
 ID AAY34173 standard; protein; 530 AA.
 XX
 AC AAY34173;
 XX
 DT 15-NOV-1999 (first entry)
 XX
 DE Human pre-proheparanase protein sequence.
 XX
 KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9943830-A2.
 XX
 PD 02-SEP-1999.
 XX
 PF 18-FEB-1999; 99WO-US001489.
 XX
 PR 24-FEB-1998; 98US-0075706P.
 PR 26-MAR-1998; 98US-0079401P.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Heinrichson RL, Fairbanks MB, Mildner AM;
 XX
 DR WPI; 1999-540598/45.
 DR N-PSDB; AAZ11236.
 XX
 PT New isolated platelet heparanase polypeptides, used to develop products
 PT for, e.g. wound healing and blocking angiogenesis.
 XX
 PS Claim 12; Fig 7; 57pp; English.
 XX
 CC This sequence is the human pre-proheparanase of the invention. This
 CC sequence was isolated from human platelets. The heparanase can be used
 CC for identifying agents which alter heparanase activity. The heparanase
 CC can be used for wound healing or for blocking angiogenesis or
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic
 CC retinopathy or solid tumours, or for the degradation of heparin and the
 CC neutralisation of heparin's anticoagulant properties during surgery.
 CC Inhibitors of heparanase activity can be used in the treatment of
 CC arthritis, asthma, and other inflammatory diseases, vascular restenosis,
 CC atherosclerosis, tumour growth and progression, fibroproliferative
 CC disorders, and central nervous system (CNS) and neurodegenerative
 CC diseases. The products can also be used for detection and diagnosis. The
 CC purified heparanase, both recombinantly produced human heparanase and
 CC heparanase isolated from human platelet activity, allows for the
 CC convenient selection of compounds having anti-heparanase activity, i.e.
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase
 CC activity. Inhibition of heparanase activity can be measured by blocking
 CC heparanase-mediated release of radioactive fragments from in vivo
 CC radiolabelled (HSPG)/heparin
 XX
 SQ Sequence 530 AA;
 Query Match 100.0%; Score 61; DB 2; Length 530;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 Db 421 CTNTDNPYK 430

RESULT 31
 AAY02345
 ID AAY02345 standard; protein; 543 AA.
 XX
 AC AAY02345;
 XX
 DT 09-JUL-1999 (first entry)

Db 424 CTNTDNPYK 433

RESULT 30
 AAY17083
 ID AAY17083 standard; protein; 532 AA.
 XX
 AC AAY17083;
 XX
 DT 21-JUL-1999 (first entry)
 XX
 DE Seq ID No: 15 of WO9921975.
 XX
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
 XX human; HSPG.
 OS Homo sapiens.
 XX
 PN WO9921975-A1.
 XX
 PD 06-MAY-1999.
 XX
 PF 28-OCT-1998; 98WO-AU000898.
 XX
 PR 28-OCT-1997; 97AU-00000062.
 PR 09-DEC-1997; 97AU-00000812.
 XX
 PA (AUSU) UNIV AUSTRALIAN NAT.
 XX
 PI Freeman CG, Hullett MD, Parish CR, Hamdorf BJ;
 XX
 DR WPI; 1999-312956/26.
 DR N-PSDB; AAX37260.
 XX
 PT Polynucleotides encoding mammalian endoglucuronidases, especially
 PT heparanases, useful to promote wound healing.
 XX
 PS Claim 6; Page 76-79; 112pp; English.
 XX
 CC The invention relates to nucleic acid sequences that encode heparanase
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are
 CC capable of removing the HS side chain from heparan sulfate proteoglycan
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
 CC inhibit heparanase, this is useful for treatment of a physiological or
 CC medical condition associated with elevated heparanase activity, such as
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
 CC rat heparanases can be used to enhance wound healing, especially
 CC associated with tissue development and repair. The conditions mentioned
 CC above can be diagnosed using specific antibodies, and also using primers
 CC and probes specific for the heparanase polynucleotides. Other uses of the
 CC heparanases include sequencing sulfated molecules such as HSPG
 XX
 SQ Sequence 532 AA;
 Query Match 100.0%; Score 61; DB 2; Length 532;
 Best Local Similarity 100.0%; Pred. No. 0.082;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 Db 437 CTNTDNPYK 446

RESULT 31
 AAY02345
 ID AAY02345 standard; protein; 543 AA.
 XX
 AC AAY02345;
 XX
 DT 09-JUL-1999 (first entry)

XX DE A human heparanase protein.

XX KW Heparanase; hp; modulator; heparin-binding growth factor;

XX KW cellular response; cytokine; cell interaction; plasma lipoprotein;

XX KW cellular susceptibility; infection; disintegration;

XX KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;

XX KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;

XX KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX OS Homo sapiens.

XX PN WO9911798-A1.

XX PD 11-MAR-1999.

XX PF 31-AUG-1998; 98WO-US017954.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 02-JUL-1998; 98US-00109386.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PA (FRIE/) FRIEDMAN M M.

XX PI Pecker I, Vlodavsky I, Feinstein E;

XX WPI; 1999-302255/25.

XX DR N-PSDB; AAX35648.

XX PT New human polynucleotide useful for treating angiogenesis, restenosis,

XX PT and inflammation.

XX PS Claim 6; Fig 1; 63pp; English.

XX CC The specification describes a polypeptide having heparanase (hp)

XX CC activity. The recombinant protein is used as a modulator of heparin-

XX CC binding growth factors, cellular responses to heparin-binding growth

XX CC factors and cytokines, cell interaction with plasma lipoproteins,

XX CC cellular susceptibility to viral, protozoal and bacterial infections or

XX CC disintegration of neurodegenerative plaques. Heparanase may be useful for

XX CC conditions such as wound healing, angiogenesis, restenosis,

XX CC atherosclerosis, inflammation, neurodegenerative diseases, and viral

XX CC infections. Mammalian heparanase can be used to neutralize plasma

XX CC heparin, and anti-heparanase antibodies may be applied for

XX CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and

XX CC renal failure in biopsy specimens, plasma samples, and body fluids. The

XX CC present sequence represents human heparanase

XX CC Sequence 543 AA;

Query Match 100.0%; Score 61; DB 2; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.084;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

DB 437 CTNTDNPYK 446

RESULT 32

AA17082

ID AAY17082 standard; protein; 543 AA.

XX AC AAY17082;

XX DT 21-JUL-1999 (first entry)

XX DE Human heparanase enzyme.

XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;

XX KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;

XX KW arteriosclerosis; atherosclerosis; inflammation; tissue development;

KW human; HSPG.

XX OS Homo sapiens.

XX PN WO9921975-A1.

XX PD 06-MAY-1999.

XX PF 28-OCT-1998; 98WO-AU000898.

XX PR 28-OCT-1997; 97AU-00000062.

XX PR 09-DEC-1997; 97AU-00000812.

XX PA (AUSU) UNIV AUSTRALIAN NAT.

XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

XX WPI; 1999-312956/26.

XX DR N-PSDB; AAX37259.

XX PT Polynucleotides encoding mammalian endoglucuronidases, especially

XX PT heparanases, useful to promote wound healing.

XX PS Claim 6; Page 69-73; 112pp; English.

XX CC The invention relates to nucleic acid sequences that encode heparanase

XX CC enzymes having endoglucuronidase activity. Recombinant heparanases are

XX CC capable of removing the HS side chain from heparan sulfate proteoglycan

XX CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to

XX CC inhibit heparanase, this is useful for treatment of a physiological or

XX CC medical condition associated with elevated heparanase activity, such as

XX CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,

XX CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

XX CC rat heparanases can be used to enhance wound healing, especially

XX CC associated with tissue development and repair. The conditions mentioned

XX CC above can be diagnosed using specific antibodies, and also using primers

XX CC and probes specific for the heparanase polynucleotides. Other uses of the

XX CC heparanases include sequencing sulfated molecules such as HSPG. The

XX CC present sequence represents a human heparanase

XX CC Sequence 543 AA;

Query Match 100.0%; Score 61; DB 2; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.084;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

DB 437 CTNTDNPYK 446

RESULT 33

AA57590

ID AAY57590 standard; protein; 543 AA.

XX AC AAY57590;

XX DT 02-MAR-2000 (first entry)

XX DE Human heparanase.

XX KW Human; heparanase; hpa; genetic modification; expression; anticancer;

XX KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;

XX KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;

XX KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;

XX KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;

XX KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;

XX KW micrometastasis; autoimmune lesion; kidney failure.

XX OS Homo sapiens.

XX PN WO9957244-A1.

PD 11-NOV-1999.
 XX 29-APR-1999; 99WO-US0009256.
 XX 01-MAY-1998; 98US-00071618.
 PR 02-MAR-1999; 99US-00260038.
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (FRIE/) FRIEDMAN M M.
 XX Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;
 PI Peleg Y, Shlomi Y;
 PI WPI; 2000-062144/05.
 DR N-PSDB; AA239195.
 XX Engineered cells that express recombinant heparanase, useful
 PT therapeutically, e.g. for treating angiogenesis and to screen for
 PT specific inhibitors, potential anticancer agents.
 XX Claim 3; Page 107-109; 118pp; English.
 XX The present invention describes genetically modified cells (A) containing
 CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
 CC and express recombinant heparanase (II). Heparanase cleaves heparan
 CC sulphate (HS) at specific intrachain sites, resulting in release of
 CC heparin-binding growth factors, enzymes and proteins that are sequestered
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It
 CC may also be implicated in tumour angiogenesis and metastases. (II) is
 CC potentially useful in wound healing and for treating angiogenesis,
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
 CC infection and cystic fibrosis. It can also be used to neutralise heparin
 CC (an alternative to protamine) and to screen for specific inhibitors
 CC (potentially useful for treating cancer and metastases). Antibodies
 CC raised against (II) are used for immunodetection and diagnosis of
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
 CC in large quantities, in a form that is homogeneously processed and
 CC activated/neutralised by a dedicated protease. The present sequence
 CC represents human heparanase
 XX Sequence 543 AA;
 SQ Query Match 100.0%; Score 61; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPRYK 10
 Db |||||
 437 CTNTDNPRYK 446
 RESULT 34
 ID AAB08849
 AC AAB08849 standard; protein; 543 AA.
 AC AAB08849;
 XX 15-JAN-2001 (first entry)
 DT Amino acid sequence of a human heparanase polypeptide.
 DE Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
 KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
 KW wound healing; infection; burn; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease;
 KW Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.
 XX Homo sapiens.
 OS WO200052178-A1.
 XX 08-SEP-2000.
 PD (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PF 14-FEB-2000; 2000WO-US003542.
 XX 01-MAR-1999; 99US-00258892.
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX Pecker I, Vlodavsky I, Feinstein E;
 XX WPI; 2000-579289/54.
 DR N-PSDB; AA475051.
 XX New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
 XX Claim 22; Fig 1; 152pp; English.
 XX The present sequence represents a human protein with heparanase catalytic
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
 CC particularly in treating tumour, inflammation or autoimmunity.
 CC Particularly, the polynucleotide is useful in modulating the
 CC bioavailability of heparin-binding growth factors, cellular responses to
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
 CC susceptibility to certain viral and some bacterial and protozoa
 CC infections, or disintegration of neurodegenerative plaques. The
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
 CC radiation burns), and in the treatment of angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
 CC bacterial or protozoa infections
 XX Sequence 543 AA;
 SQ Query Match 100.0%; Score 61; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPRYK 10
 Db |||||
 437 CTNTDNPRYK 446
 RESULT 35
 ID AAY52990
 AC AAY52990 standard; protein; 543 AA.
 AC AAY52990;
 XX 21-FEB-2000 (first entry)
 DT Human heparanase protein sequence.
 DE Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
 KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
 KW autoimmune disease; anticancer; kidney disease.
 XX Homo sapiens.
 OS WO9957153-A1.
 XX 11-NOV-1999.
 PD 29-APR-1999; 99WO-US009255.
 PF 01-MAY-1998; 98US-00071739.
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
 XX
 DR WPI; 2000-052944/04.
 DR N-PSDB; AA233290.
 XX
 XX Heparanase-specific molecular probes useful for diagnosis and treatment,
 PT e.g. of tumors, and for targeted drug delivery.
 XX
 XX Example; Page 81-82; 90pp; English.
 PS
 XX
 CC The present invention describes heparanase-specific molecular probes,
 CC useful for methods of detecting heparanase in situ. The probes and anti-
 CC heparanase antibodies are used to detect or quantify the expression of
 CC heparanase, for diagnosis and monitoring of diseases (especially
 CC metastasis), for treatment of heparanase-associated diseases (e.g.
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer for its
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
 CC sepsis and inflammatory or autoimmune disease), for targeted drug
 CC delivery (e.g. of anticancer agents) and as research reagents. The
 CC present sequence represents human heparanase, which is used in the
 CC exemplification of the present invention
 XX
 XX Sequence 543 AA;
 SQ

Query Match 100.0%; Score 61; DB 3; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 DB 437 CTNTDNPYK 446
 |||||

RESULT 36
 AAY97635
 ID AAY97635 standard; protein; 543 AA.
 AC
 AC AAY97635;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE Human heparanase protein sequence.
 XX
 KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
 KW gene therapy; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200100643-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 19-JUN-2000; 2000WO-IL000358.
 XX
 PR 25-JUN-1999; 99US-0140801P.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Pecker I, Michal I, Itzhaki H;
 XX
 DR WPI; 2001-137930/14.
 XX
 PT New polynucleotides and polypeptides that are distantly homologous to
 PT heparanase, useful in wound healing, as well as in gene therapy protocols
 PT for angiogenesis, restenosis, atherosclerosis, or inflammation.

XX Disclosure; Page 64-65; 67pp; English.
 XX
 CC This sequence represents a heparanase of the invention. The heparanase
 CC DNA and protein sequences are useful in wound healing, angiogenesis,
 CC restenosis, atherosclerosis, inflammation, pulmonary diseases,
 CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and
 CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
 CC sequence is particularly useful in gene therapy
 XX
 XX Sequence 543 AA;
 SQ

Query Match 100.0%; Score 61; DB 4; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 DB 437 CTNTDNPYK 446
 |||||

RESULT 37
 AAB86206
 ID AAB86206 standard; protein; 543 AA.
 XX
 AC AAB86206;
 XX
 DT 24-AUG-2001 (first entry)
 XX
 DE Human heparanase inhibitor protein.
 XX
 KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
 KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
 KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
 KW hepatic congestion; hydrothorax; ascite; nocturia; human.
 XX
 OS Homo sapiens.
 XX
 PN DE19955803-A1.
 XX
 PD 23-MAY-2001.
 XX
 PF 19-NOV-1999; 99DE-01055803.
 XX
 PR 19-NOV-1999; 99DE-01055803.
 XX
 PA (KNOL) KNOLL AG.
 XX
 PI Herr D, Hahn A, Laux V;
 XX
 DR WPI; 2001-368371/39.
 DR N-PSDB; AAH20940.
 XX
 PT Treatment or prevention of cardiac insufficiency and related conditions,
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of
 PT heparanase inhibitor.
 XX
 XX Disclosure; Page 11-13; 16pp; German.
 XX
 CC This invention describes a novel heparanase inhibitor which can be used
 CC for the treatment or prevention of cardiac insufficiency and associated
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
 CC the invention has cardiant, nephrotropic and hepatotropic activity. The
 CC products of the invention can be used in human and veterinary medicine,
 CC for the treatment or prevention of congestive heart failure e.g. primary
 CC cardiomyopathy. Associated conditions treated or prevented with the
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
 CC nocturia can also be treated. This sequence represents the human
 CC heparanase protein described in the method of the invention
 XX
 XX Sequence 543 AA;
 SQ

Query Match 100.0%; Score 61; DB 4; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

RESULT 38
 AAB88361
 ID AAB88361 standard; protein; 543 AA.
 XX
 AC AAB88361;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human membrane or secretory protein clone PSEC0090.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes.
 XX
 OS Homo sapiens.
 XX
 PN EP1067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2008EP-00114090.
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 DR WPI; 2001-093989/11.
 DR N-PSDB; AAF93788.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development.
 XX
 PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes

QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

Sequence 543 AA;

Query Match 100.0%; Score 61; DB 4; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

RESULT 39
 ABB07813
 ID ABB07813 standard; protein; 543 AA.
 XX
 AC ABB07813;
 XX
 DT 03-JUL-2002 (first entry)
 XX
 DE Human heparanase sequence.
 XX
 KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; human.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..35
 FT /note= "signal peptide"
 FT Protein 36..543
 FT /note= "mature protein"
 XX
 PN US20202034810-A1.
 XX
 PD 21-MAR-2002.
 XX
 PF 16-AUG-2001; 2001US-00930218.
 XX
 PR 20-SEP-2000; 2000US-00666390.
 XX
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia B;
 DR WPI; 2002-338926/37.
 XX
 PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.
 XX
 PS Disclosure; Fig 1a; 39pp; English.
 XX
 CC The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoan and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a human heparanase protein sequence used in similarity studies

QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

Query Match 100.0%; Score 61; DB 5; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 40
ADD18950
ID ADD18950 standard; protein; 543 AA.
XX AC ADD18950;
XX DT 15-JAN-2004 (first entry)
XX AC
XX DE Human disease related protein SeqID439.
XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
XX KW antiarteriosclerotic; vulnerary; gene therapy;
XX KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
XX KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
XX KW glucose transportation; catecholamine synthesis; iron transport;
XX KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
XX KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
XX KW inflammatory condition; wound healing.
XX OS Homo sapiens.
XX PN W02003018621-A2.
XX PD 06-MAR-2003.
XX PF 23-AUG-2002; 2002WO-GB003892.
XX PR 23-AUG-2001; 2001GB-00020558.
XX PR 05-OCT-2001; 2001GB-00024037.
XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.
XX PI Kingman SM., White J, Ward NR, Harris RA, Naylor S, Mundy CR;
XX DR WPI; 2003-290046/28.
XX DR N-PSDB; ADD18951.
XX PT New substantially purified polypeptide, useful for diagnosing or treating
XX PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
XX PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
XX PT wound healing.
XX PS Claim 25; SEQ ID NO 439; 424pp; English.
XX CC This invention relates to novel human genes and gene product which are
XX CC implicated in certain disease states. Compounds which modulate the
XX CC proteins of the invention may have cytostatic, antiinflammatory, the
XX CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
XX CC sequences of the invention may be useful for gene therapy. The invention
XX CC may be useful for diagnosing or treating a hypoxia-regulated condition,
XX CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
XX CC erythropoiesis, or the biological response to hypoxia conditions
XX CC including processes such as glycolysis, gluconeogenesis, glucose
XX CC transportation, catecholamine synthesis, iron transport or nitric oxide
XX CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
XX CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
XX CC inflammatory conditions or wound healing. The present sequence is that of
XX CC a disease related protein of the invention.
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
Db |||||
437 CTNTDNPYK 446

RESULT 41
ADG88800
ID ADG88800 standard; protein; 543 AA.
XX AC ADG88800;
XX DT 11-MAR-2004 (first entry)
XX AC
XX DE Human hpa protein.
XX KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
XX KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.
XX OS Homo sapiens.
XX PN US2003161823-A1.
XX PD 28-AUG-2003.
XX PF 14-JAN-2003; 2003US-00341582.
XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 05-SEP-2001; 2001WO-IL000830.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ILAN/) ILAN N.
XX PA (VLOD/) VLODAVSKY I.
XX PA (YACO/) YACOBY-ZEEVI O.
XX PA (PECK/) PECKER I.
XX PA (FEIN/) FEINSTEIN E.
XX PI Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
XX DR WPI; 2003-897910/82.
XX DR N-PSDB; ADG88799, ADG88801, ADG88832.
XX PT Composition for treating a wound comprising recombinant heparanase is
XX PT useful to induce or accelerate wound healing and induce or accelerate
XX PT angiogenesis.
XX PS Claim 2; SEQ ID NO 10; 143pp; English.
XX CC The present invention relates to methods and compositions for inducing
XX CC and/or accelerating wound healing via the catalytic activity of
XX CC heparanase. The invention is used to induce or accelerate a healing
XX CC process, particularly of an ulcer, burn, laceration, surgical incision,
XX CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate
XX CC angiogenesis. The present sequence is human hpa protein.
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
Db |||||
437 CTNTDNPYK 446

RESULT 42
ADL16379
ID ADL16379 standard; protein; 543 AA.
XX AC ADL16379;
XX DT 06-MAY-2004 (first entry)
XX AC
XX DE Human heparanase partial protein.
XX KW Human; heparanase; heparanase-dependent cancer; cancer;
XX KW autoimmune reaction; inflammation; chromosome 4; enzyme.
XX OS Homo sapiens.

```

XX US2003236215-A1.
 XX 25-DEC-2003.
 XX 09-JUN-2003; 2003US-00456573.
 XX 31-AUG-1998; 98WO-US017954.
 PR 01-MAR-1999; 99US-00258892.
 PR 08-NOV-1999; 99US-00435739.
 XX (INST-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX Pecker I, Vlodavsky I, Feinstein E;
 XX WPI; 2004-070610/07.
 XX New antisense oligonucleotide hybridizable with a polynucleotide encoding
 PT a polypeptide with heparanase activity, useful for treating diseases such
 PT as cancer and autoimmune disorders.
 XX Claim 3; SEQ ID NO 10; 108pp; English.
 XX The invention relates to an antisense oligonucleotide (ASO) comprising a
 CC polynucleotide or a polynucleotide analogue of at least 10 bases being
 CC hybridisable in vivo, under physiological conditions, with a portion of
 CC a polynucleotide strand encoding a polypeptide having heparanase
 CC catalytic activity. Also included are a method of in vivo downregulating
 CC heparanase activity (comprising administering the ASO in vivo), a method
 CC of treating a subject suffering from a pathological condition
 CC (characterised by heparanase activity, comprising administering ASO to
 CC the subject), a pharmaceutical composition comprising the ASO and a
 CC carrier, an antisense nucleic acid construct (comprising a promoter
 CC sequence and a polynucleotide sequence directing the synthesis of an
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,
 CC under physiological conditions, with a polynucleotide strand encoding a
 CC polypeptide having heparanase catalytic activity), a method of in vivo
 CC downregulating heparanase activity (comprising administering in vivo the
 CC antisense nucleic acid construct), a pharmaceutical composition
 CC comprising the antisense nucleic acid construct and a carrier, and an
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
 CC analogue of at least 10 bases being hybridisable in vivo, under
 CC physiological conditions, with a portion of a polynucleotide strand being
 CC characterised by forming at least a portion of an untranslated region
 CC (UTR) for a polynucleotide strand encoding a polypeptide having
 CC heparanase catalytic activity. The methods and compositions of the
 CC present invention are useful for the prevention and/or treatment of
 CC diseases or conditions associated with aberrant heparanase activity, such
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and
 CC inflammation. The gene for human heparanase is located on chromosome 4.
 CC The present sequence is a human heparanase protein.
 XX Sequence 543 AA;
 SQ Query Match 100.0%; Score 61; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

RESULT 43
 ADK52086
 ID ADK52086 standard; protein; 543 AA.
 XX AC ADK52086;
 XX 20-MAY-2004 (first entry)
 DT Human atopic dermatitis/psoriasis-associated protein #1.
 DE

XX Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;
 KW antipsoriatic; rash.
 XX OS Homo sapiens.
 XX WO2004016785-A1.
 XX 26-FEB-2004.
 XX 06-AUG-2003; 2003WO-JP009999.
 XX 06-AUG-2002; 2002JP-00229319.
 PR 14-MAY-2003; 2003JP-00136544.
 XX (GENO-) GENOX RES INC.
 PA (UJUU-) UNIV JUNTENDO.
 XX Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
 PI Mitsuishi K;
 XX WPI; 2004-214514/20.
 DR N-PSDB; ADK51968.
 XX Detecting atopic dermatitis or psoriasis comprises assaying levels of
 PT expression of an indicator gene at a rash site and non-rash site of a
 PT person with atopic dermatitis or psoriasis.
 XX Example 2; SEQ ID NO 119; 484pp; Japanese.
 XX The invention relates to detecting atopic dermatitis or psoriasis
 CC comprising assaying the levels of expression of an indicator gene at a
 CC rash site and non-rash site of a person with atopic dermatitis or
 CC psoriasis, comparing these levels with those of a healthy person, and
 CC determining that if the levels of indicators are higher or lower, then
 CC this indicates the disease. Also included are a reagent for detecting
 CC atopic dermatitis or psoriasis, a kit for screening for treatments, a
 CC transgenic non human vertebrate animal models for the diseases, an agent
 CC for inducing the diseases in mice and a DNA chip for assaying for the
 CC indicator genes. The method is used for treatment, detection and animal
 CC models for research of atopic dermatitis and psoriasis. The present
 CC sequence is a protein encoded by an indicator gene of the invention.
 XX Sequence 543 AA;
 SQ Query Match 100.0%; Score 61; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 |||||
 DB 437 CTNTDNPYK 446

RESULT 44
 ADM48716
 ID ADM48716 standard; protein; 543 AA.
 XX AC ADM48716;
 XX 03-JUN-2004 (first entry)
 DT Human hpa protein #1.
 DE Transgenic animal; heparanase; cancer; viral infection; restenosis;
 KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
 KW human.
 XX OS Homo sapiens.
 XX US2003217375-A1.
 XX 20-NOV-2003.
 PD

```

XX PF 24-FEB-2003; 2003US-00371218.
XX PF 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSHMIDT O.
XX PF Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX PI Chajek-Shaul T, Goldshmidt O;
XX DR WPI; 2004-021918/02.
XX DR N-PSDB; ADM48715, ADM48717.
XX XX New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancers, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX PS Example 1; SEQ ID NO 10; 106pp; English.
XX CC The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX SQ Sequence 543 AA;
XX Query Match 100.0%; Score 61; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 0.084;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 CTNTDNPYK 10
XX DB |||||
XX 437 CTNTDNPYK 446
XX RESULT 45
XX ADM48759
XX ID ADM48759 standard; protein; 543 AA.
XX AC ADM48759;
XX XX Human hpa protein #2.
XX DT 03-JUN-2004 (first entry)
XX DE
XX DE Human hpa protein #2.
XX XX Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX KW human.
XX XX Homo sapiens.
XX OS US2003217375-A1.
XX PN 20-NOV-2003.
XX PD 24-FEB-2003; 2003US-00371218.
XX PF
XX PF 24-FEB-2003; 2003US-00371218.
XX PF 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSHMIDT O.
XX PF Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX PI Chajek-Shaul T, Goldshmidt O;
XX DR WPI; 2004-021918/02.
XX DR N-PSDB; ADM48715, ADM48717.
XX XX New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancers, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX PS Example 1; SEQ ID NO 10; 106pp; English.
XX CC The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX SQ Sequence 543 AA;
XX Query Match 100.0%; Score 61; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 0.084;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 CTNTDNPYK 10
XX DB |||||
XX 437 CTNTDNPYK 446
XX RESULT 46
XX ADM05074
XX ID ADM05074 standard; protein; 543 AA.
XX AC ADM05074;
XX XX Antipsoriatic protein sequence #716.
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #716.
XX DE antipsoriatic; gene therapy; psoriasis; diagnosis.
XX KW Homo sapiens.
XX OS WO2004028479-A2.
XX PN 08-APR-2004.
XX PD 25-SEP-2003; 2003WO-US030907.
XX PF 25-SEP-2002; 2002US-0414006P.
XX PR (GETH ) GENENTECH INC.
XX PA

```

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX WPI; 2004-305105/28.
XX N-PSDB; ADN05073.
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX Claim 9; SEQ ID NO 1468; 3069pp; English.
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;
Query Match 100.0%; Score 61; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNTDNPYK 10
DB 437 CTNTDNPYK 446
RESULT 47
ID ADN04902 standard; protein; 543 AA.
AC ADN04902;
XX 01-JUL-2004 (first entry)
XX Antipsoriatic protein sequence #631.
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX Homo sapiens.
XX WO2004028479-A2.
XX 08-APR-2004.
XX 25-SEP-2003; 2003WO-US030907.
XX 25-SEP-2002; 2002US-0414006P.
XX (GETH) GENENTECH INC.
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX WPI; 2004-305105/28.
XX N-PSDB; ADN04901.
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX Claim 9; SEQ ID NO 1296; 3069pp; English.
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTNTDNPYK 10
DB 437 CTNTDNPYK 446
RESULT 48
ID ADO63831 standard; protein; 543 AA.
AC ADO63831;
XX 26-AUG-2004 (first entry)
XX Human heparanase mutant E378A.
XX Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein; enzyme.
XX Homo sapiens.
XX Synthetic.
XX Key Location/Qualifiers
FT Active-site /note= "Aactive site proton donor"
FT Active-site /note= "Active site nucleophile"
FT Misc-difference 378 /note= "Ala replaces wild-type Glu"
XX WO2004048558-A2.
XX 10-JUN-2004.
XX 24-NOV-2003; 2003WO-IL000989.
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodayvsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX Example 4; Page; 128pp; English.
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing

CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents a human
 CC heparanase mutant E378A created in an example of the invention which
 CC retains its heparanase catalytic activity. The present sequence is not
 CC shown in the invention, but is derived from the protein sequence of
 CC Genbank accession number AF14325 and the information provided on page
 CC 70.
 CC XX
 CC SQ Sequence 543 AA;
 Query Match 100.0%; Score 61; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 Db 437 CTNTDNPYK 446
 RESULT 49
 ADO63824
 ID ADO63824 standard; protein; 543 AA.
 XX
 XX AC ADO63824;
 XX
 XX DT 26-AUG-2004 (first entry)
 XX
 XX DE Human heparanase mutant E225A/E343A, SEQ ID:9.
 XX
 XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX FH Key Location/Qualifiers
 FT Misc-difference 225 /note= "Ala replaces wild-type Glu (active site proton
 FT donor)"
 FT
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site
 FT nucleophile)"
 XX
 XX WO2004048558-A2.
 XX
 XX PD 10-JUN-2004.
 XX
 XX PF 24-NOV-2003; 2003WO-IL000989.
 XX
 XX PR 24-NOV-2002; 2002IL-00153059.
 XX
 XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX PI Vlodayevsky I, Zecharia E, Goldshmidt O, Ilan N;
 XX WPI; 2004-450373/42.
 DR N-PSDB; ADO63818.
 XX
 XX PT New nucleic acid construct comprising heparanase-derived polypeptide,

PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
 PT thrombasthenia, or Bernard-Soulier syndrome.
 XX
 PS Claim 9; SEQ ID NO 9; 128pp; English.
 XX
 CC The invention relates to nucleic acid constructs comprising a nucleic
 CC acid encoding a heparanase-derived protein which lacks heparanase
 CC endoglycosidase catalytic activity but which retains its cell-cell and
 CC cell-matrix adhesion properties. The constructs of the invention
 CC optionally further comprise operably linked regulatory elements. The
 CC invention also relates to the heparanase-derived proteins and host cells
 CC comprising the nucleic acid constructs of the invention. The heparanase-
 CC derived proteins are especially mutants of human heparanase in which the
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and
 CC host cells are useful in preparing a tissue sealant composition for
 CC sealing injuries, reducing the loss of blood, accelerating the healing
 CC and homeostasis of an injury, accelerating blood vessel endothelium
 CC formation or the endothelialisation of vascular grafts, accelerating the
 CC adhesive activity of mammalian cells, and accelerating the adhesion and
 CC aggregation of platelets. They may also be used in the treatment of
 CC disorders associated with adhesion deficiency such as LAD (leukocyte
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
 CC cells of the invention may additionally be used to screen for modulators of
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable
 CC synthetic vascular graft comprising a tube made of a biocompatible
 CC material lined with the cells. The present sequence represents the human
 CC heparanase double mutant E225A/E343A.
 XX
 XX SQ Sequence 543 AA;
 Query Match 100.0%; Score 61; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTNTDNPYK 10
 Db 437 CTNTDNPYK 446
 RESULT 50
 ADO63823
 ID ADO63823 standard; protein; 543 AA.
 XX
 XX AC ADO63823;
 XX
 XX DT 26-AUG-2004 (first entry)
 XX
 XX DE Human heparanase mutant E343A, SEQ ID:8.
 XX
 XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
 KW vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnery; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX FH Key Location/Qualifiers
 FT Active-site 225 /note= "Active site proton donor"
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site
 FT nucleophile)"
 XX
 XX PN WO2004048558-A2.

PD 10-JUN-2004.
XX
XX 24-NOV-2003; 2003WO-IL000989.
XX
XX 24-NOV-2002; 2002IL-00153059.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX
XX WPI; 2004-450373/42.
DR N-PSDB; ADO63817.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX
XX Claim 9; SEQ ID NO 8; 128pp; English.
XX
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be use in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents the human
CC heparanase mutant E343A.
XX
XX Sequence 543 AA;

Query Match 100.0%; Score 61; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10
| | | | | | | | | |
Db 437 CTNTDNPYK 446

Search completed: June 5, 2006, 12:42:54
Job time : 96.589 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 11.0959 Seconds
(without alignments)
86.714 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CTNTDNPYK 10

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

PIR 80.*

1: Pirl.*

2: Pirl.*

3: Pirl.*

4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	70.5	868	2 A83547	probable aconitate
2	43	70.5	868	2 B82213	aconitate hydratase
3	43	70.5	868	2 C81200	aconitate hydratase
4	43	70.5	868	2 H81775	aconitate hydratase
5	41	67.2	87	2 S01982	hypothetical prote
6	40	65.6	297	2 F97218	dihydrodipicolinat
7	39	63.9	259	2 F97218	glycosyltransferas
8	38	62.3	384	2 T44652	UDP-N-acetylglucos
9	38	62.3	487	2 T49424	hypothetical prote
10	38	62.3	616	2 T07611	aconitate hydratase
11	38	62.3	848	2 G86708	aconitate hydratase
12	38	62.3	855	2 C30262	aconitate hydratase
13	38	62.3	870	2 E72541	probable aconitate
14	38	62.3	878	2 A97854	aconitate hydratase
15	38	62.3	878	2 A71641	aconitate hydratase
16	38	62.3	887	2 T27868	hypothetical prote
17	38	62.3	889	1 A44153	aconitate hydratase
18	38	62.3	889	2 A44154	aconitate hydratase
19	38	62.3	889	2 S26403	aconitate hydratase
20	38	62.3	889	2 S18720	aconitate hydratase
21	38	62.3	890	2 A00270	aconitate hydratase
22	38	62.3	891	2 A00654	aconitate hydratase
23	38	62.3	891	2 B48642	aconitate hydratase
24	38	62.3	891	2 A90860	aconitate hydratase
25	38	62.3	891	2 E85759	aconitate hydratase
26	38	62.3	891	2 G84875	aconitate hydratase
27	38	62.3	895	2 A13483	aconitate hydratase
28	38	62.3	897	2 AD2906	aconitate hydratase
29	38	62.3	897	2 E97681	aconitate hydratase

30	38	62.3	898	2 B84471	cytoplasmic aconit
31	38	62.3	898	2 T04693	aconitate hydratase
32	38	62.3	898	2 T10101	aconitate hydratase
33	38	62.3	900	2 A11279	aconitate hydratase
34	38	62.3	900	2 A11642	aconitate hydratase
35	38	62.3	901	2 F89910	aconitate hydratase
36	38	62.3	903	2 A87704	aconitate hydratase
37	38	62.3	906	2 G75362	aconitate hydratase
38	38	62.3	907	2 T04820	aconitate hydratase
39	38	62.3	908	2 G82824	aconitate hydratase
40	38	62.3	909	2 G69599	aconitate hydratase
41	38	62.3	910	2 B83451	aconitate hydratase
42	38	62.3	943	2 F70873	aconitate hydratase
43	38	62.3	944	2 G87135	aconitate hydratase
44	37	60.7	201	2 A12205	hypothetical prote
45	37	60.7	323	2 A96670	hypothetical prote
46	37	60.7	661	2 I52603	MPS1 protein - mou
47	37	60.7	756	2 F64112	malate dehydrogena
48	37	60.7	805	2 E88412	protein C44P1.5 li
49	37	60.7	805	2 T19936	hypothetical prote
50	37	60.7	1260	2 T37523	probable oxoprolin
51	37	60.7	1470	2 S45323	genome polyprotein
52	36	59.0	151	2 P00506	hypothetical prote
53	36	59.0	604	2 D90523	lipoprotein (impor
54	36	59.0	728	2 H65208	hypothetical 82.6
55	36	59.0	728	2 E86093	regulator of acety
56	36	59.0	728	2 G91245	probable gamma-ada
57	36	59.0	865	2 T41685	plasmacytoma-asso
58	36	59.0	1028	2 A53449	BIG-1 protein - ra
59	36	59.0	1028	2 I58164	hypothetical prote
60	35	57.4	117	2 F72536	MHC class I RT1.C-
61	35	57.4	343	2 I69009	hypothetical prote
62	35	57.4	384	2 T19265	probable glutamate
63	35	57.4	402	2 T37879	levansucrase - Brw
64	35	57.4	415	2 S39195	glutamate 5-kinase
65	35	57.4	423	2 S46741	glutamate 5-kinase
66	35	57.4	428	2 S61186	C4b-binding protei
67	35	57.4	610	1 I46001	F14010.2 protein -
68	35	57.4	739	2 F86337	hypothetical prote
69	35	57.4	813	2 T02672	DNA topoisomerase
70	35	57.4	868	2 G64119	probable permease
71	35	57.4	875	2 B97236	iodide peroxidase
72	35	57.4	914	1 JN0550	iodide peroxidase
73	35	57.4	914	1 S07047	iodide peroxidase
74	35	57.4	915	2 A49874	metabotropic gluta
75	35	57.4	926	1 OPPGIT	iodide peroxidase
76	35	57.4	933	1 OPHUIT	iodide peroxidase
77	35	57.4	1215	2 E70614	hypothetical prote
78	34	55.7	72	2 G97134	hypothetical prote
79	34	55.7	98	2 H34964	Ig heavy chain V-I
80	34	55.7	98	2 A49051	Ig heavy chain V-I
81	34	55.7	117	2 S18554	Ig heavy chain V r
82	34	55.7	120	2 S26789	Ig heavy chain V r
83	34	55.7	131	2 S26792	Ig heavy chain V r
84	34	55.7	136	2 T48077	hypothetical prote
85	34	55.7	275	1 B64077	bis(5'-nucleosyl)-
86	34	55.7	289	2 B86649	hypothetical prote
87	34	55.7	321	2 H69869	transcription regu
88	34	55.7	341	2 T16951	hypothetical prote
89	34	55.7	374	2 S07845	mRNA maturase b13
90	34	55.7	429	2 T16656	hypothetical prote
91	34	55.7	469	2 S44620	C50C3.1 protein -
92	34	55.7	485	2 C71400	adenosylhomocyste
93	34	55.7	517	2 S21209	mRNA maturase b13
94	34	55.7	517	2 S78664	mRNA maturase b13
95	34	55.7	599	1 A54906	afamin precursor -
96	34	55.7	682	2 T12968	hypothetical prote
97	34	55.7	770	2 T23999	hypothetical prote
98	34	55.7	1074	2 T24877	hypothetical prote
99	34	55.7	1076	2 T24887	hypothetical prote
100	34	55.7	1402	2 T24664	hypothetical prote
101	34	55.7	1437	2 S07430	M polyprotein prec
102	34	55.7	2135	2 T14602	variant-specific s

103	33.5	54.9	680	2	D64420	N-methylhydantoina	176	32	52.5	173	2	AD0718	probable lipoprote
104	33.5	54.9	1002	2	T30546	major surface glyc	177	32	52.5	179	2	BH3830	peptide methionine
105	33.5	54.9	1162	2	B97852	hypothetical prote	178	32	52.5	184	2	H90592	hypothetical prote
106	33.5	54.9	1169	2	T71639	hypothetical prote	179	32	52.5	205	2	T15757	hypothetical prote
107	33	54.1	103	2	F83918	hypothetical prote	180	32	52.5	229	2	S44217	coat protein - tom
108	33	54.1	103	2	T21100	hypothetical prote	181	32	52.5	241	1	QOCVW2	coat protein - abu
109	33	54.1	164	2	G71427	hypothetical prote	182	32	52.5	241	2	G90190	conserved hypotet
110	33	54.1	179	2	PN0096	hypothetical 19K p	183	32	52.5	244	2	JC8019	CD58 protein - pig
111	33	54.1	181	2	A31245	outer spore coat p	184	32	52.5	247	1	QOCVR1	coat protein - tom
112	33	54.1	185	2	G64088	transcription anti	185	32	52.5	251	1	JQ1869	coat protein - tom
113	33	54.1	187	1	G70406	siroheme synthase	186	32	52.5	251	1	QOCVP2	coat protein - pot
114	33	54.1	224	1	A46343	nonstructural prot	187	32	52.5	251	1	QOCVS2	coat protein - pep
115	33	54.1	224	1	JQ0597	nonstructural prot	188	32	52.5	251	2	QOCVS2	coat protein - pep
116	33	54.1	225	2	A31926	transcription acti	189	32	52.5	251	2	S11874	ARI protein - pepp
117	33	54.1	225	2	S48363	hypothetical prote	190	32	52.5	256	1	JQ1886	coat protein - tom
118	33	54.1	252	2	S08054	hypothetical prote	191	32	52.5	256	2	F97163	flagellar basal bo
119	33	54.1	277	2	D89455	protein F55A4.2 [i	192	32	52.5	256	2	JQ2326	coat protein - Ind
120	33	54.1	290	2	T37151	probable DNA-bindi	193	32	52.5	256	2	S22589	hypothetical prote
121	33	54.1	330	2	I69645	probable nucleotid	194	32	52.5	256	2	S58346	coat protein - cass
122	33	54.1	334	2	F82038	adenosine deaminas	195	32	52.5	257	2	JQ2332	ARI protein - cass
123	33	54.1	350	2	T39555	hypothetical prote	196	32	52.5	257	2	S39210	gene V2 protein -
124	33	54.1	372	2	A83167	conserved hypotet	197	32	52.5	257	2	S39234	gene V2 protein -
125	33	54.1	393	1	A48573	calreticulin autoa	198	32	52.5	258	1	QOCMCI	coat protein - cas
126	33	54.1	395	2	D85570	hypothetical prote	199	32	52.5	258	1	VCOMCN	coat protein - cas
127	33	54.1	428	2	G81504	enolase CP1071 [im	200	32	52.5	258	2	S25624	coat protein - cas
128	33	54.1	428	2	A72034	enolase - Chlamydo	201	32	52.5	258	2	S25622	coat protein - cas
129	33	54.1	428	2	R66590	enolase [imported]	202	32	52.5	258	2	S25623	coat protein - cas
130	33	54.1	458	2	A12722	tail fiber protein	203	32	52.5	260	1	QOCVCL	coat protein - tom
131	33	54.1	459	2	T08713	NAD+ ADP-ribosyltr	204	32	52.5	263	2	T16329	hypothetical prote
132	33	54.1	462	2	E97504	gene 17, tail fibre	205	32	52.5	280	2	D82097	methionine aminope
133	33	54.1	485	2	T06764	adenosylhomocyste	206	32	52.5	285	2	AC1537	hypothetical prote
134	33	54.1	504	2	I57022	integral membrane	207	32	52.5	316	2	E84636	NAM (no apical mer
135	33	54.1	521	2	S02018	regulatory protein	208	32	52.5	325	2	F82350	lipid A biosynthes
136	33	54.1	535	2	C95040	glucan 1,6-alpha-9	209	32	52.5	325	2	A71164	hypothetical prote
137	33	54.1	537	2	F90465	medium-chain-fatty	210	32	52.5	332	2	H82340	gluconate utilizac
138	33	54.1	583	2	T20291	hypothetical prote	211	32	52.5	335	2	T31730	hypothetical prote
139	33	54.1	604	2	H89914	hypothetical prote	212	32	52.5	335	2	T32657	hypothetical prote
140	33	54.1	611	2	A54086	calnexin-t - mouse	213	32	52.5	341	2	D75068	glutamine-fructose
141	33	54.1	611	2	A53418	calnexin precursor	214	32	52.5	388	2	S48879	protein kinase SMK
142	33	54.1	629	2	T39285	probable transmemb	215	32	52.5	393	2	S06259	major outer membra
143	33	54.1	641	2	T52489	Rec10 protein [val	216	32	52.5	394	2	S11012	major outer membra
144	33	54.1	652	2	S25265	outer membrane pro	217	32	52.5	395	2	F86740	teichoic acid bios
145	33	54.1	652	2	D82317	iron-regulated out	218	32	52.5	399	2	B83398	hypothetical prote
146	33	54.1	692	2	A97013	hypothetical prote	219	32	52.5	408	2	T16601	hypothetical prote
147	33	54.1	718	2	A51832	ATP-dependent DNA	220	32	52.5	428	2	S09134	gene ND4L intron 1
148	33	54.1	770	2	S60676	cellobiose oxidase	221	32	52.5	429	2	T15303	hypothetical prote
149	33	54.1	791	2	S55725	rec10 protein - fi	222	32	52.5	439	2	T30787	hypothetical prote
150	33	54.1	803	2	F90485	hypothetical prote	223	32	52.5	439	2	T28472	hypothetical prote
151	33	54.1	837	2	A83383	hypothetical membr	224	32	52.5	439	2	H72154	hypothetical prote - var
152	33	54.1	872	2	H86435	protein F17F8.5 [i	225	32	52.5	439	2	F42507	F10L protein - vac
153	33	54.1	889	2	C72585	probable valyl-tRN	226	32	52.5	439	2	E36840	C14L protein - var
154	33	54.1	982	2	E88465	protein B0244.6 [i	227	32	52.5	446	2	T00846	hypothetical prote
155	33	54.1	983	2	E86989	probable integral	228	32	52.5	470	2	S6942	hypothetical prote
156	33	54.1	991	2	T48631	polynucleotide pho	229	32	52.5	472	2	A12997	transcription regu
157	33	54.1	1037	2	E81980	pilus-associated p	230	32	52.5	482	2	H98285	probable transcrip
158	33	54.1	1050	2	JC7889	heparinase (EC 3.2	231	32	52.5	492	2	T21934	hypothetical prote
159	33	54.1	1060	2	T43046	retrovirus-related	232	32	52.5	510	2	E82374	threonine ammonia-
160	33	54.1	1150	2	T13824	LK6 protein kinase	233	32	52.5	514	2	T40485	transmembrane tran
161	33	54.1	1260	2	A86323	protein F14D16.3 [234	32	52.5	527	2	D87318	conserved hypotet
162	33	54.1	1399	2	A99720	RhaC core protein	235	32	52.5	532	2	T52442	hypothetical prote
163	33	54.1	1401	2	T02255	probable ubiquitou	236	32	52.5	538	2	I68093	PRR2 delta - human
164	33	54.1	1558	2	T29233	hypothetical prote	237	32	52.5	606	2	T13152	WDRI protein - hum
165	33	54.1	2133	2	T30637	hypothetical prote	238	32	52.5	607	1	ABXL72	74K albumin precur
166	33	54.1	2514	1	MNWV82	nonstructural poly	239	32	52.5	615	1	ABCHS	serum albumin prec
167	33	54.1	3144	2	A46068	Huntington disease	240	32	52.5	620	2	T30765	hypothetical prote
168	33	54.1	8243	2	T31307	type I fatty acid	241	32	52.5	624	1	JC4510	pullulanase (EC 3.
169	32.5	53.3	443	2	C83421	hypothetical prote	242	32	52.5	630	2	D97992	hypothetical prote
170	32	52.5	133	2	E30587	outer membrane pro	243	32	52.5	631	2	S72270	alpha-amylase (EC
171	32	52.5	133	2	F30587	outer membrane pro	244	32	52.5	632	2	A95923	hypothetical prote
172	32	52.5	154	2	H83160	hypothetical prote	245	32	52.5	678	2	A71287	probable cytoplasm
173	32	52.5	155	2	C60333	outer membrane pro	246	32	52.5	758	2	I64084	hemoglobin recepto
174	32	52.5	167	2	C64456	hypothetical prote	247	32	52.5	770	2	A49283	outer layer protei
175	32	52.5	168	1	S19361	hypothetical prote	248	32	52.5	773	2	T01539	hypothetical prote

395	31	50.8	686	2	A59348	formate dehydrogen	468	30	49.2	170	2	T50215	probable peptide m
396	31	50.8	687	2	A41905	ferric vibriobacti	469	30	49.2	183	2	A81821	DNA-3-methyladenin
397	31	50.8	692	2	T47493	hypothetical prote	470	30	49.2	183	2	C81057	DNA-3-methyladenin
398	31	50.8	716	2	E89998	conserved hypotet	471	30	49.2	184	2	S34250	27.2K structural p
399	31	50.8	725	1	E64211	virulence-associat	472	30	49.2	187	2	S62511	probable peptide m
400	31	50.8	769	1	IJHULM	leukocyte adhesio	473	30	49.2	204	2	T19249	hypothetical prote
401	31	50.8	769	1	JC1121	leukocyte adhesio	474	30	49.2	206	2	JL0059	H-2 class I histoc
402	31	50.8	770	2	S04847	leukocyte adhesio	475	30	49.2	206	2	JL0058	H-2 class I histoc
403	31	50.8	771	2	A45839	leukocyte adhesio	476	30	49.2	211	2	E71334	hypothetical prote
404	31	50.8	776	2	T29064	hyaluronate lyase	477	30	49.2	216	2	A61259	glycoprotein S - p
405	31	50.8	785	2	T09491	hemagglutinin, pha	478	30	49.2	218	2	T49885	peptide methionine
406	31	50.8	797	2	AH1302	primosomal replica	479	30	49.2	219	2	T17539	glycerophosphoryl
407	31	50.8	797	2	AH1674	primosomal replica	480	30	49.2	221	2	A57296	ribosomal protein
408	31	50.8	807	2	D86350	DNA topoisomerase	481	30	49.2	238	2	B64404	hypothetical prote
409	31	50.8	813	2	A13479	FBK7.13 protein -	482	30	49.2	249	2	T19088	hypothetical prote
410	31	50.8	862	2	F75116	hypothetical prote	483	30	49.2	249	2	T131837	hypothetical prote
411	31	50.8	879	2	A47704	endoglucanase I (E	484	30	49.2	253	2	A36564	hemoglobin linker
412	31	50.8	882	2	AB1631	valyl-tRNA synthet	485	30	49.2	261	2	A86763	cobryic acid synth
413	31	50.8	899	2	I38153	gene ret11 protein	486	30	49.2	261	2	T07630	expansin 1 - tomat
414	31	50.8	937	2	A45082	neurotrophic recep	487	30	49.2	299	2	T26789	hypothetical prote
415	31	50.8	940	2	S19702	fibronectin-bindin	488	30	49.2	299	2	S54087	probable membrane
416	31	50.8	961	2	G90053	hypothetical prote	489	30	49.2	307	2	B72031	lipic acid synthe
417	31	50.8	987	2	A64474	hypothetical prote	490	30	49.2	307	2	F86594	lipote synthetase
418	31	50.8	1038	2	H90053	hypothetical prote	491	30	49.2	309	2	A38395	mast cell carboxyp
419	31	50.8	1050	2	S45636	natriuretic-peptid	492	30	49.2	314	2	A00220	flagellar protein
420	31	50.8	1051	2	T18351	lmp1 protein - Myc	493	30	49.2	314	2	AF3193	conserved hypotet
421	31	50.8	1097	2	T40678	hypothetical prote	494	30	49.2	319	2	B35090	MHC nonclassical c
422	31	50.8	1114	1	S05582	protein-tyrosine k	495	30	49.2	319	2	D64303	hypothetical prote
423	31	50.8	1120	2	H71664	transcription-repa	496	30	49.2	326	2	A32273	MHC class I histoc
424	31	50.8	1127	2	T25804	hypothetical prote	497	30	49.2	328	2	I54414	MHC H-2K transplan
425	31	50.8	1300	2	S73679	probable lipoprote	498	30	49.2	328	2	S51458	hypothetical prote
426	31	50.8	1301	2	S18118	alpha-amylase - Al	499	30	49.2	330	2	AD2258	hypothetical prote
427	31	50.8	1302	1	JC6009	surface-located me	500	30	49.2	330	2	AC2240	hypothetical prote
428	31	50.8	1347	2	T02214	ubiquitous TPR mot	501	30	49.2	332	2	G96840	hypothetical prote
429	31	50.8	1353	2	T26301	hypothetical prote	502	30	49.2	347	2	T32768	hypothetical prote
430	31	50.8	1365	2	T30822	lmp1 protein - Myc	503	30	49.2	347	2	E84898	hypothetical prote
431	31	50.8	1475	2	T29809	hypothetical prote	504	30	49.2	348	2	A83984	C4-dicarboxylate t
432	31	50.8	1513	1	RNR202	DNA-directed RNA p	505	30	49.2	351	2	C91182	probable fibmrial
433	31	50.8	1638	2	T25352	hypothetical prote	506	30	49.2	351	2	T44428	probable gonococca
434	31	50.8	1722	2	E89753	protein fl1C7.4 li	507	30	49.2	351	2	G86028	probable fibmrial
435	31	50.8	1827	1	A23945	sucrose alpha-gluc	508	30	49.2	351	2	A23351	serendipity (ary)
436	31	50.8	1854	2	S36859	c1pA protein - Clo	509	30	49.2	353	2	S25336	MHC class I histoc
437	31	50.8	2215	2	T16871	hypothetical prote	510	30	49.2	356	2	A21198	H-2 class I histoc
438	31	50.8	2481	2	A43908	fibronectin - Afri	511	30	49.2	356	2	T18590	hypothetical prote
439	31	50.8	2581	2	AP2545	hypothetical prote	512	30	49.2	358	2	S75659	gene At103 protein
440	31	50.8	4544	1	S02352	alpha-2-macroglobu	513	30	49.2	360	2	I54554	MHC class I histoc
441	31	50.8	4545	1	S25111	alpha-2-macroglobu	514	30	49.2	362	1	HLMSDB	MHC class I histoc
442	31	50.8	4976	2	T14165	peptide synthetase	515	30	49.2	362	2	C60854	MHC class I histoc
443	30.5	50.0	454	2	S75741	hypothetical prote	516	30	49.2	362	2	G08054	MHC class I histoc
444	30.5	50.0	523	2	T36968	probable phytoene	517	30	49.2	362	1	B10304	protein-tyrosine-p
445	30.5	50.0	2183	1	GNNY84	genome polypotein	518	30	49.2	364	1	S31304	MHC class I-alpha
446	30	49.2	53	2	JS0105	hypothetical 5.8K	519	30	49.2	365	2	I57814	MHC class I histoc
447	30	49.2	55	2	C82814	hypothetical prote	520	30	49.2	367	2	S31651	MHC class I histoc
448	30	49.2	91	2	I54404	T cell antigen rec	521	30	49.2	368	1	HLMSKD	MHC class I Rtl.E
449	30	49.2	98	2	S03497	T-cell receptor be	522	30	49.2	368	2	I49712	MHC class I histoc
450	30	49.2	113	2	I38312	T-cell receptor be	523	30	49.2	368	2	I49713	MHC class I histoc
451	30	49.2	116	2	T24843	hypothetical prote	524	30	49.2	368	2	I68705	protein kinase (EC
452	30	49.2	116	2	T16332	hypothetical prote	525	30	49.2	369	1	HLMSKB	protein kinase (EC
453	30	49.2	116	2	H90966	hypothetical prote	526	30	49.2	369	1	HLMSKB	protein kinase (EC
454	30	49.2	119	2	H90513	hypothetical prote	527	30	49.2	370	2	I54531	protein kinase (EC
455	30	49.2	129	1	QC0VW4	AC2 protein - abut	528	30	49.2	373	2	I69008	protein kinase (EC
456	30	49.2	129	2	JG0019	flagellar basal-bo	529	30	49.2	377	2	A45851	MHC class I histoc
457	30	49.2	134	2	A30563	T-cell receptor be	530	30	49.2	379	1	OKHUR1	protein kinase (EC
458	30	49.2	137	2	T15522	hypothetical prote	531	30	49.2	379	1	OKB01R	protein kinase (EC
459	30	49.2	140	2	S48562	hypothetical prote	532	30	49.2	380	1	OKPGLR	protein kinase (EC
460	30	49.2	142	2	S19245	Ig heavy chain pre	533	30	49.2	381	1	A60669	protein kinase (EC
461	30	49.2	142	4	S13768	MHC class I histoc	534	30	49.2	381	1	OKHULR	protein kinase (EC
462	30	49.2	145	2	S65983	vybn protein - Bac	535	30	49.2	381	1	OKMSR1	protein kinase (EC
463	30	49.2	148	2	AC1897	hypothetical prote	536	30	49.2	381	1	OKRT1R	protein kinase (EC
464	30	49.2	161	2	S74336	hypothetical prote	537	30	49.2	385	2	T18180	proline-rich prote
465	30	49.2	161	2	E89812	conserved hypotet	538	30	49.2	392	2	A06777	keratin 2, type I,
466	30	49.2	164	2	T16321	hypothetical prote	539	30	49.2	395	1	A25887	calsequestrin prec
467	30	49.2	165	2	S252511	hypothetical prote	540	30	49.2	395	1	A46345	gene Iii protein -

541 30 49.2 406 1 A31050
542 30 49.2 407 2 S50870
543 30 49.2 408 2 T24114
544 30 49.2 409 1 JQ1396
545 30 49.2 410 1 A28071
546 30 49.2 414 2 T50010
547 30 49.2 420 1 S22418
548 30 49.2 421 2 S67132
549 30 49.2 423 2 S38953
550 30 49.2 429 2 E75107
551 30 49.2 431 2 F71600
552 30 49.2 438 2 A64428
553 30 49.2 441 2 T11715
554 30 49.2 442 2 T42091
555 30 49.2 448 2 T45710
556 30 49.2 448 2 C86447
557 30 49.2 451 1 J66180
558 30 49.2 452 2 S10840
559 30 49.2 457 2 S63608
560 30 49.2 463 2 G84901
561 30 49.2 464 2 A28879
562 30 49.2 464 2 A97656
563 30 49.2 468 1 P2ML51
564 30 49.2 470 2 H69296
565 30 49.2 480 2 A80148
566 30 49.2 480 2 S72668
567 30 49.2 480 2 JC7866
568 30 49.2 484 2 E64432
569 30 49.2 494 2 S57539
570 30 49.2 506 2 AE2475
571 30 49.2 506 2 B52210
572 30 49.2 507 2 T38653
573 30 49.2 510 2 E37307
574 30 49.2 513 2 T38770
575 30 49.2 521 2 T34583
576 30 49.2 521 2 E82377
577 30 49.2 522 2 B82024
578 30 49.2 522 2 G81243
579 30 49.2 530 2 JN0597
580 30 49.2 531 2 B88109
581 30 49.2 532 2 T49873
582 30 49.2 534 2 A29776
583 30 49.2 537 2 T48599
584 30 49.2 540 2 T10892
585 30 49.2 541 2 T46423
586 30 49.2 546 2 T06415
587 30 49.2 547 2 S54513
588 30 49.2 548 2 A70980
589 30 49.2 555 1 B69862
590 30 49.2 558 2 S68447
591 30 49.2 561 2 E30891
592 30 49.2 561 2 D85726
593 30 49.2 561 2 C64903
594 30 49.2 566 2 T33042
595 30 49.2 567 2 G90537
596 30 49.2 571 2 G83226
597 30 49.2 572 2 S77249
598 30 49.2 586 2 T19075
599 30 49.2 587 2 A60367
600 30 49.2 595 2 T06412
601 30 49.2 602 1 JQ0806
602 30 49.2 604 1 TVYUH
603 30 49.2 608 2 S65298
604 30 49.2 617 2 S08317
605 30 49.2 619 2 E83635
606 30 49.2 620 2 T39074
607 30 49.2 636 2 E50943
608 30 49.2 636 2 A85792
609 30 49.2 636 2 H64941
610 30 49.2 637 2 F96900
611 30 49.2 637 2 T46555
612 30 49.2 639 2 JQ0607
613 30 49.2 639 2 JT0479

calsequestrin prec
cell division cont
hypothetical prote
calsequestrin prec
calsequestrin prec
hypothetical prote
calsequestrin prec
probable membrane
carboxypeptidase D
thiamin biosynthes
variant-specific s
thiamin biosynthes
probable acetylorn
acetylornithine tr
H-protein promoter
F5D14.3 protein -
stearoyl-CoA 9-des
gene ND4L intron p
glutamate dehydrog
hypothetical prote
two component sens
probable two-compo
L2 protein - human
acetyl-CoA decarbo
catalase (EC 1.11.
oviduct-specific p
endo-1,6-beta-D-gl
spore coat polysac
probable membrane
polyketide synthas
hetM protein - Ana
trp-asp repeat pro
probable cardiolip
alpha-amylose a pr
probable periplasm
ABC transporter, p
peptide methionine
peptide methionine
calnexin-like prot
protein T24E12.5 [
calnexin homolog -
glucan 1,4-alpha-g
hypothetical prote
probable calnexin
hypothetical prote
calnexin - soybean
probable membrane
probable accB5 pro
conserved hypothet
origin recognition
hypothetical prote
hypothetical prote
probable ABC-type
hypothetical prote
lipoprotein limpor
probable two-compo
hypothetical prote
hypothetical prote
transforming prote
probable chaperoni
hydrogenase (EC 1.
protein-tyrosine k
dicarboxylic amino
hydrogenase (EC 1.
hypothetical prote
yeast ches homolog
probable enzyme li
probable enzyme Z2
Probable ATP-depen
DNA gyrase (topois
DNA topoisomerase
glucan 1,4-alpha-g

614 615
616 616
617 618
618 619
619 620
620 621
621 622
622 623
623 624
624 625
625 626
626 627
627 628
628 629
629 630
630 631
631 632
632 633
633 634
634 635
635 636
636 637
637 638
638 639
639 640
640 641
641 642
642 643
643 644
644 645
645 646
646 647
647 648
648 649
649 650
650 651
651 652
652 653
653 654
654 655
655 656
656 657
657 658
658 659
659 660
660 661
661 662
662 663
663 664
664 665
665 666
666 667
667 668
668 669
669 670
670 671
671 672
672 673
673 674
674 675
675 676
676 677
677 678
678 679
679 680
680 681
681 682
682 683
683 684
684 685
685 686

640 1 ALASGR
640 2 A29166
640 2 A64065
647 2 A49218
648 2 T28325
649 2 S48752
655 1 A46688
667 2 AD0295
677 2 G84614
685 2 S64158
696 2 T46111
698 1 TVFVLV
704 2 JC7881
726 1 S73915
742 1 S58691
808 2 A82877
826 2 B36203
861 2 T41945
875 2 T33434
877 2 H64708
879 2 E71811
880 1 SYBSVS
902 2 T47966
903 2 B84548
907 2 T27317
916 2 F71962
925 2 T00781
930 2 T42241
938 2 T25215
941 1 TVMVMD
952 2 B57238
963 2 A57238
967 2 G86229
982 2 A97210
985 2 E69850
1001 2 S66704
1004 2 T71490
1011 2 T17430
1018 2 A32192
1025 2 H81751
1042 2 G64514
1042 2 T41146
1077 2 I46497
1099 2 I46498
1099 2 I46496
1155 2 S64365
1174 1 HJBYDH
1180 1 NCECX5
1180 2 G85933
1180 2 E91088
1181 2 AB0865
1205 2 T18517
1205 2 T18517
1223 1 TVCHLV
1224 2 T00059
1254 1 JQ1978
1254 1 VHWVVE
1254 1 VHWVVT
1254 1 B44213
1255 1 D44213
1255 2 T31344
1289 2 T43060
1328 2 T47492
1334 2 AB1775
1348 2 S27812
1348 2 A43917
1348 2 F84486
1356 2 T47492
1363 2 T47492
1403 1 A47328
1447 1 VGIHE3
1447 1 VGIHE2
1447 1 A43573
1449 1 VGIHFS
1449 2 S47423
1451 1 JQ1719

glucan 1,4-alpha-g
glucan 1,4-alpha-g
hypothetical prote
hemagglutinin homo
ORF MSV164 probabl
major surface prot
hepatocyte growth
probable formyl tr
hypothetical prote
hypothetical prote
probable transposa
protein-tyrosine k
ceramidase (EC 3.5
virulence-associat
kinesin-related pr
hypothetical prote
iron-regulated out
probable iron-regu
valine-tRNA ligase
hypothetical prote
hypothetical prote
myotubularin prote
hypothetical prote
protein-tyrosine k
iron-responsive el
iron-responsive el
hypothetical prote
beta galactosidase
formate dehydrogen
hypothetical prote
probable exodeoxyr
col protein - Neur
fibronectin-bindin
exodeoxyribonuclea
type I restriction
probable cysteine-
bumentanide-sensiti
bumentanide-sensiti
GDP/GTP exchange p
helicase (EC 3'6.1
exodeoxyribonuclea
DNA helicase RecB
DNA helicase RecB
exonuclease V chai
procollagen N-endo
epidermal growth f
hypothetical prote
structural polypro
structural polypro
structural polypro
structural polypro
structural polypro
Gp80 precursor - s
agrin - electric r
hypothetical prote
probable epidermal
probable epidermal
copla-like polypro
natural killer cel
E2 glycoprotein pr
E2 glycoprotein pr
E2 glycoprotein pr
E2 glycoprotein pr
E2 glycoprotein pr

687	30	49.2	1452	1	VGIH79	E2 glycoprotein pr	760	29	47.5	268	2	D86270	F21P23.14 protein
688	30	49.2	1453	2	S41453	spike protein - ca	761	29	47.5	268	2	T00827	hypothetical prote
689	30	49.2	1454	3	T18560	DNA-directed DNA p	762	29	47.5	269	2	AE3455	arginine-binding p
690	30	49.2	1538	2	T29095	cardiac muscle fac	763	29	47.5	271	2	F64349	hypothetical prote
691	30	49.2	1680	2	T41628	probable transcrip	764	29	47.5	271	2	S12560	transcription fact
692	30	49.2	1785	2	A45546	major merozoite su	765	29	47.5	272	2	G71638	merozoite surface
693	30	49.2	1856	2	C95008	immunoglobulin A1	766	29	47.5	274	2	B90007	protein F59B1.5 [i
694	30	49.2	2195	2	T34264	hypothetical prote	767	29	47.5	276	2	I40453	licheninase (EC 3.
695	30	49.2	2287	2	T21312	hypothetical prote	768	29	47.5	285	2	C97279	thioredoxin reduct
696	30	49.2	2496	2	A71616	secreted protein p	769	29	47.5	285	2	B82842	spermidine synthas
697	30	49.2	5825	2	T12117	polyprotein - fava	770	29	47.5	287	2	T40138	probable ribosomal
698	30	49.2	6359	2	T13679	bacitracin synthet	771	29	47.5	289	2	S69801	L-alanoyl-D-glutam
699	29.5	48.4	226	2	T22710	hypothetical prote	772	29	47.5	289	2	A11448	L-alanoyl-D-glutam
700	29.5	48.4	435	2	C86340	protein F2D10.28 [773	29	47.5	291	2	H97090	thioredoxin reduct
701	29.5	48.4	648	2	AF0609	conserved hypoteth	774	29	47.5	293	1	DDBP32	helix-destabilizin
702	29.5	48.4	800	1	S31575	interleukin-4 rece	775	29	47.5	293	1	DDBP36	helix-destabilizin
703	29.5	48.4	899	2	H87513	hypothetical prote	776	29	47.5	294	2	G83962	hypothetical prote
704	29	47.5	29	2	B54197	70K thyroid autoan	777	29	47.5	295	2	F95997	probable transcrip
705	29	47.5	62	2	C86560	conserved hypoteth	778	29	47.5	295	4	S36174	RNA binding protei
706	29	47.5	62	2	B72064	conserved hypoteth	779	29	47.5	296	2	B82866	conjugal transfer
707	29	47.5	71	2	T00008	copy number contro	780	29	47.5	296	2	C82605	conjugal transfer
708	29	47.5	73	2	T07454	hypothetical prote	781	29	47.5	299	2	B90111	DNA-directed RNA p
709	29	47.5	84	2	JN0469	85K WRK-20 recogni	782	29	47.5	301	1	DDBP34	helix-destabilizin
710	29	47.5	96	2	T13118	secondary immunity	783	29	47.5	305	1	PRYNB	Na+/K+-exchanging
711	29	47.5	96	2	AG0229	conserved hypoteth	784	29	47.5	307	2	E89075	protein K04A8.2 [i
712	29	47.5	100	2	A46503	beta-integrin - ra	785	29	47.5	310	2	S36581	E2 protein - human
713	29	47.5	107	2	S08521	hypothetical 11.7K	786	29	47.5	310	2	AB2235	hypothetical prote
714	29	47.5	115	2	B84969	flagellar basal-bo	787	29	47.5	311	2	T49912	hypothetical prote
715	29	47.5	119	2	AF2509	hypothetical prote	788	29	47.5	312	2	S04281	psbA intron 2 prot
716	29	47.5	129	2	H70727	hypothetical prote	789	29	47.5	314	2	B90334	hypothetical prote
717	29	47.5	140	2	AC0512	probable secreted	790	29	47.5	323	2	S01895	T-cell receptor ga
718	29	47.5	141	2	H85060	hypothetical prote	791	29	47.5	323	2	T46285	hypothetical prote
719	29	47.5	143	2	AH3060	hypothetical prote	792	29	47.5	324	2	JC2395	Fas antigen precur
720	29	47.5	158	2	T49567	related to attachm	793	29	47.5	332	2	T19150	hypothetical prote
721	29	47.5	163	2	T71635	hypothetical prote	794	29	47.5	334	2	H95307	arginine deiminase
722	29	47.5	167	2	H98225	hypothetical prote	795	29	47.5	335	2	A85913	probable enzyme X3
723	29	47.5	168	2	S71562	drought-induced pr	796	29	47.5	335	2	G91068	probable enzyme EC
724	29	47.5	176	2	G81269	probable acetyltra	797	29	47.5	336	2	T10733	cinnamoyl-CoA redu
725	29	47.5	177	2	S47348	histone H1.0 - rat	798	29	47.5	336	2	A89772	hypothetical prote
726	29	47.5	180	2	B45613	surface antigen FU	799	29	47.5	337	2	A83403	ribose operon repr
727	29	47.5	184	2	G88449	protein F54D8.5 [i	800	29	47.5	339	2	H97107	uncharacterized pr
728	29	47.5	186	2	G64397	hypothetical prote	801	29	47.5	343	2	G71979	probable type II D
729	29	47.5	188	2	A29867	hypothetical 20K p	802	29	47.5	351	2	T50105	hypothetical prote
730	29	47.5	194	1	HSU010	histone H1.0 - hum	803	29	47.5	351	2	G84128	hypothetical prote
731	29	47.5	194	2	I49150	histone H1.0 - mou	804	29	47.5	352	2	AC3306	periplasmic dipept
732	29	47.5	196	1	RSXJ5A	histone H5A - Afri	805	29	47.5	354	2	S27014	GTP-binding regula
733	29	47.5	196	2	A30484	histone H5B - Afri	806	29	47.5	355	2	E81415	DNA-directed DNA p
734	29	47.5	210	2	B65045	hypothetical prote	807	29	47.5	356	2	S61061	hypothetical prote
735	29	47.5	211	2	T24760	hypothetical prote	808	29	47.5	361	2	I48160	MHC class I protei
736	29	47.5	214	2	B46244	insulin-like growt	809	29	47.5	363	2	S61970	hypothetical prote
737	29	47.5	217	2	T19005	hypothetical prote	810	29	47.5	363	2	S49539	CoD protein precu
738	29	47.5	220	2	AE3221	conserved hypoteth	811	29	47.5	365	2	C97735	hypothetical prote
739	29	47.5	224	2	E96982	ortholog yrbG, yet	812	29	47.5	369	2	S63484	hypothetical prote
740	29	47.5	224	2	G72746	hypothetical prote	813	29	47.5	371	2	T03025	mitosis-specific c
741	29	47.5	224	2	B72710	hypothetical prote	814	29	47.5	371	2	D97042	hypothetical prote
742	29	47.5	228	2	I58170	LERK-7 precursor -	815	29	47.5	373	2	S48451	probable membrane
743	29	47.5	234	2	T26363	hypothetical prote	816	29	47.5	373	2	AH0855	lipoprotein NlpD p
744	29	47.5	240	2	T38050	hypothetical prote	817	29	47.5	376	2	G71925	cag island protein
745	29	47.5	241	2	T23428	hypothetical prote	818	29	47.5	377	2	T40024	probable cytochrom
746	29	47.5	246	2	T30490	hypothetical prote	819	29	47.5	379	2	D91078	probable lipoprote
747	29	47.5	246	2	T37473	transcription regu	820	29	47.5	379	2	B85923	lipoprotein (impor
748	29	47.5	247	2	T21406	hypothetical prote	821	29	47.5	379	2	B55522	lipoprotein D prec
749	29	47.5	250	2	S66156	Rieske iron-sulfur	822	29	47.5	381	2	S37170	repB protein - Lac
750	29	47.5	251	2	T25121	hypothetical prote	823	29	47.5	383	2	B86272	protein F16A14.12
751	29	47.5	255	1	E71090	probable lactam ut	824	29	47.5	384	2	A29130	calreticulin (clon
752	29	47.5	255	2	H75074	lactam utilization	825	29	47.5	384	2	AE2024	sulfolipid biosynt
753	29	47.5	257	2	H84157	essential for sigm	826	29	47.5	388	2	S40057	repA protein - Lac
754	29	47.5	257	2	T21029	hypothetical prote	827	29	47.5	388	2	T22795	hypothetical prote
755	29	47.5	260	2	H97357	stage 0 sporulatio	828	29	47.5	392	2	S37902	hypothetical prote
756	29	47.5	260	2	S47724	hypothetical 29.5K	829	29	47.5	401	2	T17515	hypothetical prote
757	29	47.5	261	2	S20610	calpastatin - mous	830	29	47.5	402	2	E64440	hypothetical prote
758	29	47.5	262	2	T43753	probable COI intro	831	29	47.5	409	2	D83326	probable acyl-CoA
759	29	47.5	264	2	AE2274	hypothetical prote	832	29	47.5	411	2	S29129	calreticulin precu

833	29	47.5	414	2	D96838	unknown protein T2	906	29	47.5	575	2	S75395	probable glutamine
834	29	47.5	416	1	S06763	calreticulin precu	907	29	47.5	579	2	B86158	F22D16.20 protein
835	29	47.5	416	2	JH0819	calreticulin precu	908	29	47.5	582	2	H97017	ATP-dependent Zn p
836	29	47.5	417	1	A37047	calreticulin precu	909	29	47.5	582	2	A70755	hypothetical prote
837	29	47.5	418	1	A34154	calreticulin precu	910	29	47.5	585	2	C82157	hypothetical prote
838	29	47.5	419	2	S71343	calreticulin precu	911	29	47.5	586	2	S63386	HOLI protein - yea
839	29	47.5	419	2	F72866	alkaline exonuclea	912	29	47.5	589	2	JG7520	endo-1,6-alpha-D-m
840	29	47.5	419	2	E86292	FYH2.15 protein -	913	29	47.5	598	2	T28238	ORF MSV077 hypothe
841	29	47.5	420	2	T41870	ALK-EXO ofr133 - B	914	29	47.5	599	2	A11025	hypothetical prote
842	29	47.5	421	2	T25383	hypothetical prote	915	29	47.5	602	2	T13219	major capsid prote
843	29	47.5	423	2	T50923	acetylornithine tr	916	29	47.5	603	2	S15074	calpastatin - rat
844	29	47.5	423	2	S57103	hypothetical prote	917	29	47.5	607	2	A47757	retrovirus-related
845	29	47.5	425	2	E64657	probable glucose-6	918	29	47.5	609	2	A30894	70K thyroid autoan
846	29	47.5	425	2	T34872	hypothetical prote	919	29	47.5	610	2	A84417	hypothetical prote
847	29	47.5	427	2	T19004	hypothetical prote	920	29	47.5	614	2	T19963	hypothetical prote
848	29	47.5	429	2	S09141	ND4L intron 2 prot	921	29	47.5	619	2	A60646	transforming prote
849	29	47.5	429	2	S45459	TOM34 protein - ye	922	29	47.5	624	2	PC6003	surface membrane p
850	29	47.5	432	1	S18932	u-plasminogen acti	923	29	47.5	624	2	B83386	hypothetical prote
851	29	47.5	433	1	JN0560	u-plasminogen acti	924	29	47.5	634	2	AG0252	probable DEAD box
852	29	47.5	433	2	A81793	deoxyribodipyrimid	925	29	47.5	636	2	A50725	probable ATP-depen
853	29	47.5	436	1	S49458	diphosphate-fructo	926	29	47.5	644	2	S50429	dnak-type molecula
854	29	47.5	437	2	F59099	hypothetical prote	927	29	47.5	654	2	T10772	calpastatin - rat
855	29	47.5	442	1	UKPG	u-plasminogen acti	928	29	47.5	656	1	A34890	histidine decarbox
856	29	47.5	449	1	KIECD3	aspartate kinase (929	29	47.5	658	2	T03416	traG protein - Agr
857	29	47.5	449	2	G91254	lysine sensitive a	930	29	47.5	658	2	AB3243	conjugal transfer
858	29	47.5	449	2	C86095	aspartokinase III,	931	29	47.5	662	1	S12989	histidine decarbox
859	29	47.5	449	2	T08309	conserved hypotet	932	29	47.5	664	2	A97222	membrane associate
860	29	47.5	455	2	T38275	hypothetical prote	933	29	47.5	668	2	I39902	penicillin-binding
861	29	47.5	456	2	T45610	proanthranilate N-	934	29	47.5	682	1	RNRZC1	DNA-directed RNA p
862	29	47.5	457	2	C86669	amino acid permeas	935	29	47.5	684	2	T33785	hypothetical prote
863	29	47.5	458	2	A84487	probable replicati	936	29	47.5	688	2	S85241	hypothetical prote
864	29	47.5	463	2	T16503	hypothetical prote	937	29	47.5	705	2	C75118	dipeptide abc tran
865	29	47.5	464	2	B57720	keatine receptor b	938	29	47.5	719	2	AB1358	hypothetical prote
866	29	47.5	470	1	NMIV98	exo-alpha-sialidas	939	29	47.5	732	2	A43315	BTS domain protein
867	29	47.5	470	1	NMIV9	exo-alpha-sialidas	940	29	47.5	736	2	A99279	hypothetical prote
868	29	47.5	473	2	C84312	glycine dehydrogen	941	29	47.5	736	2	E71414	hypothetical prote
869	29	47.5	474	2	S30168	mercury(II) reduct	942	29	47.5	750	1	COZPME	mei2 protein - fis
870	29	47.5	474	2	A35732	protective protein	943	29	47.5	751	2	I48748	semaphorin E - mou
871	29	47.5	475	2	B75024	glutamate synthase	944	29	47.5	754	2	T16182	hypothetical prote
872	29	47.5	475	2	F70144	carboxyl-terminal	945	29	47.5	754	2	AH3004	vgrG protein (impo
873	29	47.5	476	2	D71200	probable glutamate	946	29	47.5	759	2	S62067	TV1 enhancer activ
874	29	47.5	481	2	AB1527	hypothetical prote	947	29	47.5	767	2	S41479	DNA-binding protei
875	29	47.5	489	2	S67802	hypothetical prote	948	29	47.5	775	2	T45238	probable transfera
876	29	47.5	489	2	S76768	hypothetical prote	949	29	47.5	797	2	D71621	hypothetical prote
877	29	47.5	492	2	T26502	hypothetical prote	950	29	47.5	798	2	A40526	integrin beta-7 ch
878	29	47.5	492	2	S75049	hypothetical prote	951	29	47.5	806	2	A46271	integrin beta-7 ch
879	29	47.5	502	2	AE3569	fructuronate reduc	952	29	47.5	811	2	S77577	endo-alpha-sialida
880	29	47.5	502	2	T17416	probable alkylhali	953	29	47.5	816	2	G84952	aspartate kinase (
881	29	47.5	505	2	C82216	probable fumarate	954	29	47.5	816	2	B98196	hypothetical prote
882	29	47.5	510	2	H84887	probable pectinest	955	29	47.5	816	2	AH3090	VgrG protein (impo
883	29	47.5	514	2	T26501	hypothetical prote	956	29	47.5	828	2	AH2443	hypothetical prote
884	29	47.5	517	2	A49776	xylan 1,4-beta-xy	957	29	47.5	836	2	AF3233	conserved hypotet
885	29	47.5	521	2	H82318	formate-tetrahydro	958	29	47.5	845	2	T52518	related to cytosin
886	29	47.5	526	2	A86274	F/A19.15 protein -	959	29	47.5	850	2	S45553	SIN3 protein-bindi
887	29	47.5	527	2	F83319	probable thiosulfa	960	29	47.5	851	2	AD1427	internain, probab
888	29	47.5	529	2	C82314	methyl-accepting c	961	29	47.5	853	2	H70470	hypothetical prote
889	29	47.5	529	2	S35306	phytoene dehydroge	962	29	47.5	855	2	T38754	beta-transducin -
890	29	47.5	531	2	T11596	hypothetical prote	963	29	47.5	857	2	T05172	hypothetical prote
891	29	47.5	540	1	TVFVEB	protein-tyrosine k	964	29	47.5	859	1	VCLJCT	env polyprotein pr
892	29	47.5	540	2	D86737	malolactac enzyme	965	29	47.5	863	2	T38016	importin beta-1 su
893	29	47.5	540	2	S38728	malolactac enzyme	966	29	47.5	863	2	AH2296	hypothetical prote
894	29	47.5	540	2	B44776	protein-tyrosine k	967	29	47.5	879	2	S49910	chloroplast outer
895	29	47.5	544	2	S35745	protein-tyrosine k	968	29	47.5	882	2	A39030	androgen-binding p
896	29	47.5	545	2	S00727	kinase-related tra	969	29	47.5	896	2	S26984	probable DNA-direc
897	29	47.5	546	1	VGN2RK	cell fusion glycop	970	29	47.5	903	2	T09143	alpha-glucosidase
898	29	47.5	546	2	A73305	gene F protein - r	971	29	47.5	903	2	A10015	maltoose regulon po
899	29	47.5	547	2	T25478	hypothetical prote	972	29	47.5	913	2	JC5463	alpha-glucosidase
900	29	47.5	556	2	AD11394	arginyl tRNA synth	973	29	47.5	914	2	S46593	finger protein AZP
901	29	47.5	556	2	AD1770	arginyl tRNA synth	974	29	47.5	931	2	D86222	protein F7G19.9 [i
902	29	47.5	556	2	S22634	sphingomyelin phos	975	29	47.5	938	2	A39160	transcription acti
903	29	47.5	558	2	C71609	hypothetical prote	976	29	47.5	940	2	T41992	hypothetical prote
904	29	47.5	570	2	AD2292	hypothetical prote	977	29	47.5	941	2	I40772	hypothetical prote
905	29	47.5	574	2	T33794	hypothetical prote	978	29	47.5	946	2	F81361	probable cell divi

979 29 47.5 955 2 T39765 probable nuclear m
980 29 47.5 966 2 T30017 hypothetical prote
981 29 47.5 969 2 T23256 hypothetical prote
982 29 47.5 978 1 RGBYI3 regulatory protein
983 29 47.5 984 2 H90029 hypothetical prote
984 29 47.5 989 2 C83035 hypothetical prote
985 29 47.5 996 2 JE0237 apolipoprotein E r
986 29 47.5 1004 1 S55353 probable copper-tr
987 29 47.5 1025 1 JC1266 beta-galactosidase
988 29 47.5 1047 2 T25782 hypothetical prote
989 29 47.5 1073 2 T01955 hypothetical prote
990 29 47.5 1082 1 RNEGBB DNA-directed RNA p
991 29 47.5 1093 2 A31758 phosphorylase kina
992 29 47.5 1093 2 B40793 phosphorylase kina
993 29 47.5 1097 2 JQ0301 hypothetical 127K
994 29 47.5 1141 2 A44093 cGMP-inhibited cAM
995 29 47.5 1172 2 A42587 thrombospondin 2 p
996 29 47.5 1178 1 A39804 thrombospondin pre
997 29 47.5 1208 2 T39068 coiled coil protei
998 29 47.5 1254 1 JQ1979 structural polypro
999 29 47.5 1268 2 T18955 hypothetical prote
1000 29 47.5 1280 2 T34357 hypothetical prote

ALIGNMENTS

RESULT 1
A83547
probable aconitase hydratase PA0794 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2004
C:Accession: A83547
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <STO>
A:Cross-references: UNIPROT:Q915E4; UNIPARC:UPI00000C5157; GB:AE004514; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0794
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 411 CTNTSNPR 418

RESULT 2
B82213
aconitase hydratase 1 VC1338 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 31-Dec-2004
C:Accession: B82213
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Hardison, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82213
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-868 <HEI>
A:Cross-references: UNIPROT:Q9KSC0; UNIPARC:UPI00000C2F63; GB:AE004213; GB:AE003852; NID:
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1338
A:Map position: 1
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 415 CTNTSNPR 422

RESULT 3
C81200
aconitase hydratase 1 NMB0433 [imported] - Neisseria meningitidis (strain MC58 serogroup
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Dec-2004
C:Accession: C81200
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81200
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <TET>
A:Cross-references: UNIPROT:Q9K0X3; UNIPARC:UPI00000C44A8; GB:AE002399; GB:AE002098; NID:
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0433
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 412 CTNTSNPR 419

RESULT 4
H81775
aconitase hydratase (EC 4.2.1.3) NMA2052 [imported] - Neisseria meningitidis (strain Z2491)
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 31-Dec-2004
C:Accession: H81775
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: H81775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-868 <PAR>
A:Cross-references: UNIPROT:Q9J705; UNIPARC:UPI00000332C3; GB:AL162758; GB:AL157959; NID:
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: acnA; NMA2052
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 13;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8

Db 412 CTNTSNPR 419

RESULT 5

S01982

hypothetical protein 87 - phase T5

C:Species: phase T5

C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 09-Jul-2004

C:Accession: S01982

R:Kalinan, A.V.; Kryukov, V.M.; Bayev, A.A.

Nucleic Acids Res. 16, 6230, 1988

A:Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early and

A:Reference number: S01982; PMID:88289370; PMID:3267228

A:Accession: S01982

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-87 <KAL>

A:Cross-references: UNIPROT:P13390; UNIPARC:UPI000017A852; EMBL:X07559

C:Genetics:

A:Start codon: GTG

Query Match 67.2%; Score 41; DB 2; Length 87;

Best Local Similarity 70.0%; Pred. No. 3.2;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10

Db 61 CTNTDNLQK 70

RESULT 6

D86827

dihydrodipicolinate synthase (EC 4.2.1.52) [imported] - Lactococcus lactis subsp. lactis

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004

C:Accession: D86827

R:Bohloian, A.; Winkler, P.; Mager, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A:Reference number: A86625; PMID:21235186; PMID:11337471

A:Accession: D86827

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-297 <STO>

A:Cross-references: UNIPROT:Q9CF61; UNIPARC:UPI000012854B; GB:AE005176; PID:gi2724628; E

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: dapA

C:Superfamily: dihydrodipicolinate synthase

C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 65.6%; Score 40; DB 2; Length 297;

Best Local Similarity 77.8%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNPY 9

Db 174 CTNTDNLAY 182

RESULT 7

F97218

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: F97218

R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4836, 2001

Query Match 62.3%; Score 38; DB 2; Length 487;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: F97218

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <KUR>

A:Cross-references: UNIPROT:Q97FY6; UNIPARC:UPI00000CA596; GB:AE001437; PIDN:AAK80537.1;

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3588

C:Superfamily: dolichyl-phosphate beta-D-mannosyltransferase

Query Match 63.9%; Score 39; DB 2; Length 259;

Best Local Similarity 66.7%; Pred. No. 20;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNPY 9

Db 103 CDNTQDPY 111

RESULT 8

T44652

UDP-N-acetylglucosamine-2-epimerase cpsO [imported] - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004

C:Accession: T44652

R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nittayajarn, A.; Rubens, C.E.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z22821

A:Accession: T44652

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-384 <CHA>

A:Cross-references: UNIPROT:Q9RPB9; UNIPARC:UPI00000B43B7; EMBL:AF163833; PIDN:AAD53075.1

A:Experimental source: strain COH1; serotype III

C:Genetics:

A:Gene: cpsO

Query Match 62.3%; Score 38; DB 2; Length 384;

Best Local Similarity 77.8%; Pred. No. 44;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TINTDNPYK 10

Db 342 TINTDNPYK 350

RESULT 9

T49424

hypothetical protein B17C10.50 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004

C:Accession: T49424

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49424

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-487 <SCH>

A:Cross-references: UNIPROT:Q9P6E4; UNIPARC:UPI000006AEB4; EMBL:AL355926; GSPDB:GN00116;

A:Experimental source: BAC clone B17C10; strain OR74A

C:Genetics:

A:Gene: NCSP:B17C10.50

A:Map position: 6

A:Introns: 86/1; 150/2

C:Superfamily: Neurospora crassa hypothetical protein B17C10.50

Query Match 62.3%; Score 38; DB 2; Length 487;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
Db 150 CTGTDNP 156

RESULT 10
T07611
aconitate hydratase (EC 4.2.1.3) - potato (fragment)
N;Alternate names: aconitase
C;Species: Solanum tuberosum (potato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Dec-2004
C;Accession: T07611
R;Surpili, M.J.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z16049
A;Accession: T07611
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-616 <SUR>
A;Cross-references: UNIPROT:O04916; UNIPARC:UPI00001252D0; EMBL:X97012; PIDN:CAA65735.1
A;Experimental source: cv. Desiree
C;Function:
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also to
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: carbon-oxygen lyase, hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 616;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
Db 159 CTNTSNP 165

RESULT 11
G86708
aconitate hydratase (EC 4.2.1.3) [imported] - Lactococcus lactis subsp. lactis (strain I
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Dec-2004
C;Accession: G86708
R;Solotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: G86708
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-848 <STO>
A;Cross-references: UNIPROT:Q9CHQ5; UNIPARC:UPI00000C68A0; GB:AE005176; PID:gl2723578; F
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: citB
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: carbon-oxygen lyase, hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 848;
Best Local Similarity 85.7%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
Db 400 CTNTSNP 406

RESULT 12
C90262
aconitate hydratase [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 31-Dec-2004
C;Accession: C90262
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: C90262
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-855 <KUR>
A;Cross-references: UNIPROT:Q97242; UNIPARC:UPI0000064353; GB:AE006641; NID:gl13814284; P
C;Genetics:
A;Gene: SS01095
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 855;
Best Local Similarity 85.7%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
Db 402 CTNTSNP 408

RESULT 13
E72541
probable aconitate hydratase APE1618 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 31-Dec-2004
C;Accession: E72541
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: E72541
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-870 <KAW>
A;Cross-references: UNIPROT:Q9VB10; UNIPARC:UPI000005E007; DDBJ:AP000062; NID:gs105244; I
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1618
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 870;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
Db 406 CTNTSNP 412

RESULT 14
A97854
aconitate hydratase (EC 4.2.1.3) [imported] - Rickettsia conorii (strain Malish 7)
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 31-Dec-2004
C;Accession: A97854
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID:21442074; PMID:11557893
A;Accession: A97854
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-878 <KUR>
A;Cross-references: UNIPROT:Q92G90; UNIPARC:UPI00001252DE; GB:AE006914; PIDN:AAL03771.1;
C;Genetics:
A;Gene: acnA
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: carbon-oxygen lyase, hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 878;

Best Local Similarity 85.7%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
|||||
Db 426 CTNTSNP 432

RESULT 15
A71641
aconitate hydratase (acna) RP799 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 31-Dec-2004
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71641; MUID:99039499; PMID:9823893
A:Accession: A71641
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-878 <AND>
A:Cross-references: UNIPROT:Q92CF4; UNIPARC:UPI00001252DF; GB:AJ235269; NID
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: acna; RP799
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 878;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
|||||
Db 426 CTNTSNP 432

RESULT 16
T27868
hypothetical protein ZK455.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Dec-2004
C:Accession: T27868
R:White, S.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z20432
A:Accession: T27868
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-887 <WIL>
A:Cross-references: UNIPROT:Q23500; UNIPARC:UPI00001252CD; EMBL:Z66567; PIDN:CAA91491.1;
A:Experimental source: clone ZK455
C:Genetics:
A:Gene: CESP:ZK455.1
A:Map position: X
A:Introns: 31/1; 87/2; 133/2; 156/3; 426/2; 522/3; 715/3
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 887;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
|||||
Db 436 CTNTSNP 442

RESULT 17
A44153
aconitate hydratase (EC 4.2.1.3) - rabbit
N:Alternate names: aconitase; ferritin mRNA repressor protein; iron-responsive element-b
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Dec-2004

C:Accession: A44153; S47664; S47665
R:Patino, M.M.; Walden, W.E.
J. Biol. Chem. 267, 19011-19016, 1992
A:Title: Cloning of a functional cDNA for the rabbit ferritin mRNA repressor protein. Den
A:Reference number: A44153; MUID:92406828; PMID:1527028
A:Accession: A44153
A:Molecule type: mRNA; protein
A:Residues: 1-889 <PAT>
A:Cross-references: UNIPROT:Q01059; UNIPARC:UPI000012D87F; EMBL:M95815; NID:gl65029; PID
A:Experimental source: liver
A>Note: sequence extracted from NCBI backbone (NCBIN:113547, NCBIP:113546)
R:Swenson, G.R.; Walden, W.E.
Nucleic Acids Res. 22, 2627-2633, 1994
A:Title: Localization of an RNA binding element of the iron responsive element binding pr
A:Reference number: S47664; MUID:94316507; PMID:7518918
A:Accession: S47664
A:Molecule type: protein
A:Residues: 133,'X',135-146 <SWE>
A:Cross-references: UNIPARC:UPI00000172FE8
A:Accession: S47665
A:Molecule type: protein
A:Residues: 624-635,'X',637-638 <SWM>
A:Cross-references: UNIPARC:UPI0000172FE8
C:Genetics:
A:Gene: FRP
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein
F:480-623/Domain: RNA binding #status predicted <RNA>
F:125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pred
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 1; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

RESULT 18
A44154
aconitate hydratase (EC 4.2.1.3) - rat
N:Alternate names: aconitase; iron-responsive element-binding protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2004
C:Accession: A44154
R:Yu, Y.; Radisky, E.; Leibold, E.A.
J. Biol. Chem. 267, 19005-19010, 1992
A:Title: The iron-responsive element binding protein. Purification, cloning, and regulati
A:Reference number: A44154; MUID:92406827; PMID:1527027
A:Accession: A44154
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-889 <YUI>
A:Cross-references: UNIPARC:UPI00000176090
A:Experimental source: liver
A>Note: sequence extracted from NCBI backbone (NCBIN:113545, NCBIP:113544)
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein
F:480-623/Domain: RNA binding #status predicted <RNA>
F:125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pred
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

```

RESULT 19
S26403
acconitate hydratase (EC 4.2.1.13) - human
N:Alternate names: aconitase; iron regulatory factor; iron-responsive element-binding protein
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C:Accession: S26403; A36203; S22147
R:Hirling, H.; Emery-Goodman, A.; Thompson, N.; Neupert, B.; Seiser, C.; Kuehn, L.C.
Nucleic Acids Res. 20, 33-39, 1992
A:Title: Expression of active iron regulatory factor from a full-length human cDNA by in
A:Reference number: S26403; MUID:92150156; PMID:1738601
A:Accession: S26403
A:Molecule type: mRNA
A:Residues: 1-889 <HIR>
A:Cross-references: UNIPROT:P21399; UNIPARC:UPI000012D87E; EMBL:Z11559; NID:g33962; PIDN
R:Rouault, T.A.; Tang, C.K.; Kaptain, S.; Burgees, W.H.; Haile, D.J.; Samaniego, F.; McB
Proc. Natl. Acad. Sci. U.S.A. 87, 7958-7962, 1990
A:Title: Cloning of the cDNA encoding an RNA regulatory protein--the human iron-respons
A:Reference number: A36203; MUID:91045916; PMID:2172968
A:Accession: A36203
A:Molecule type: mRNA
A:Residues: 100-531, 'II', 534, 'TGILKAELY', 544-711, 'PEVWTPSWHGEHLPTL', 728-889 <ROU>
A:Cross-references: UNIPARC:UPI000017608F; GB:M37836
C:Genetics:
A:Gene: GDB:IREB1
A:Cross-references: GDB:125344; OMIM:147581
A:Map position: 9pter-9qter
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)
C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
F:480-623/Domain: RNA binding #status predicted <RNA>
F:125,126,178,205,207,302,778/Active site: Asp, His, Asp, His, Glu, Ser #status pre
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

RESULT 20
S18720
acconitate hydratase (EC 4.2.1.13) - mouse
N:Alternate names: aconitase; iron-responsive element-binding protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 31-Dec-2004
C:Accession: S18720
R:Philpott, C.C.; Rouault, T.A.; Klausner, R.D.
Nucleic Acids Res. 19, 6333, 1991
A:Title: Sequence and expression of the murine iron-responsive element binding protein.
A:Reference number: S18720; MUID:92066494; PMID:1956798
A:Accession: S18720
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-889 <PHI>
A:Cross-references: UNIPROT:P28271; UNIPARC:UPI000002805A; EMBL:X61147; NID:g52735; PIDN
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1991
C:Genetics:
A:Gene: irebP; acol
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)
C:Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur pro
F:480-623/Domain: RNA binding #status predicted <RNA>
F:125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

```

```

Db 437 CTNTSNP 443

RESULT 21
AH0270
acconitate hydratase (EC 4.2.1.13) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 31-Dec-2004
C:Accession: AH0270
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0270
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-890 <KUR>
A:Cross-references: UNIPROT:Q8ZEFL; UNIPARC:UPI00000DC909; GB:AL590842; PIDN:CAC91028.1;
C:Genetics:
A:Gene: acnA
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 890;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 22
AH0654
acconitate hydratase 1 (citrate hydro-lyase 1) [imported] - Salmonella enterica subsp. ent
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 31-Dec-2004
C:Accession: AH0654
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conneron, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0654
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-891 <PAR>
A:Cross-references: UNIPARC:UPI0000059F3C; GB:AL513382; PIDN:CAD08419.1; PID:g16502462;
C:Genetics:
A:Gene: STY1339
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 23
B48642
acconitate hydratase (EC 4.2.1.13) - Legionella pneumophila
C:Species: Legionella pneumophila
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 31-Dec-2004
C:Accession: B48642
R:Mengaud, J.M.; Horwitz, M.A.

```

J. Bacteriol. 175, 5666-5676, 1993
A;Title: The major iron-containing protein of Legionella pneumophila is an aconitase hom
A;Reference number: A48642; MUID:93374864; PMID:8366052
A;Accession: B48642
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-891 <MEN>
A;Cross-references: UNIPROT:P37032; UNIPARC:UPI00001252DB; GB:L22081; NID:G348943; PIDN:
A;Experimental source: strain Philadelphia 1, substrain pneumophila
C;Genetics:
A;Gene: acn
C;Function:
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t
A;Pathway: tricarboxylic acid cycle
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
F;478-620/Domain: RNA binding #status predicted <RNA>
F;130,131,183,214,216,311,776/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre
F;435,501,504/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 24
Aconitase hydratase 1 [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 31-Dec-2004
C;Accession: A90860
R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A90860
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-891 <HAY>
A;Cross-references: UNIPROT:Q8X7C7; UNIPARC:UPI00000DOA85; GB:BA0000007; PIDN:BA035272.1;
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: Ecbl849
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 25
E85759
Aconitase hydratase 1 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Dec-2004
C;Accession: E85759
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E85759
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-891 <STO>
A;Cross-references: UNIPROT:Q8X7C7; UNIPARC:UPI00000DOA85; GB:AE005174; NID:G12515516; P
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: acnA
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 26
G64875
Aconitase hydratase (EC 4.2.1.3) - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 31-Dec-2004
C;Accession: G64875; S22375; A49756
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64875
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-891 <BLAT>
A;Cross-references: UNIPROT:P25516; UNIPARC:UPI00000480C9; GB:AE0000225; GB:U00096; NID:G
A;Experimental source: strain K-12, substrain MG1655
R;Prodromou, C.; Artymuk, P.J.; Guest, J.R.
Eur. J. Biochem. 204, 599-609, 1992
A;Title: The aconitase of Escherichia coli. Nucleotide sequence of the aconitase gene and
isopropylmalate isomerases.
A;Reference number: S22374; MUID:92174916; PMID:1541275
A;Accession: S22375
A;Molecule type: DNA
A;Residues: 1-521, 'G', 523-891 <PRO>
A;Cross-references: UNIPARC:UPI000016ED57; EMBL:X60293; NID:G40894; PIDN:CAA42834.1; PID:
A;Experimental source: strain K-12, substrain W3110
R;Prodromou, C.; Haynes, M.J.; Guest, J.R.
J. Gen. Microbiol. 137, 2505-2515, 1991
A;Title: The aconitase of Escherichia coli: purification of the enzyme and molecular clor
A;Reference number: A49756; MUID:92148368; PMID:1838390
A;Accession: A49756
A;Molecule type: protein
A;Residues: 'X', 3-15, 'X', 17, 'X', 19 <PR2>
A;Cross-references: UNIPARC:UPI0000176094
C;Genetics:
A;Gene: acnA; acn
C;Function:
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t
A;Pathway: tricarboxylic acid cycle
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
F;2-891/Product: aconitate hydratase #status predicted <MA>
F;478-620/Domain: RNA binding #status predicted <RNA>
F;128,129,181,212,214,309,776/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre
F;435,501,504/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 27

A:Molecule type: mRNA

A:Residues: 'NFLUSSENRSLYFASSLDLYLSS', 1-129, 'Y', 131-231, 'RP', 234-714, 'VAVVMRLWREH', 728-8

A:Cross-references: UNIPARC:UPI000011D718; EMBL:X82839; NID:G599624; PID:G599625

A:Experimental source: variety ecotype Columbia; unripened pods

A>Note: the extension at the 5' end results from translation of the 5' untranslated regi

A>Note: the differences near the carboxyl end are due to a frameshift error

C:Genetics:

A:Gene: ACO

A:Map position: 4

A:Introns: 2/1; 37/1; 90/3; 121/3; 183/3; 237/3; 286/1; 333/3; 352/3; 379/2; 404/3; 466/

A>Note: F4B14.100

C:Function:

A:Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also c

A:Pathway: glyoxylate bypass

C:Superfamily: aconitate A/homoaconitase (aconitate hydratase 1)

C:Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur pro

F:484-627/Domains: RNA binding #status predicted <RNA>

F:441,507,510/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 898;

Best Local Similarity 85.7%; Pred. NO. 1e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7

Db 441 CTNTSNP 447

RESULT 32

T10101

aconitate hydratase (EC 4.2.1.3) - cucurbit

N:Alternate names: aconitase

C:Species: Cucurbita sp. (cucurbit)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 31-Dec-2004

C:Accession: T10101

R:Hayashi, M.; De Bellis, L.; Alpi, A.; Nishimura, M.

Plant Cell Physiol. 36, 669-680, 1995

A:Title: Cytosolic aconitase participates in the glyoxylate cycle in etiolated pumpkin c

A:Reference number: Z16948; MUID:95368272; PMID:7640891

A:Accession: T10101

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-898 <HAY>

A:Cross-references: UNIPARC:UPI00001252CE; EMBL:D29629; NID:G868002; PIDN:BAA06108.1; PI

A:Experimental source: cv. Kurokawa Amakuri Nankin; etiolated cotyledons

C:Function:

A:Description: catalyzes the reversible hydration of cis-aconitate to citrate and also t

A:Pathway: tricarboxylic acid cycle

C:Superfamily: aconitate A/homoaconitase (aconitate hydratase 1)

C:Keywords: carbon-oxygen lyase; cytosol; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 898;

Best Local Similarity 85.7%; Pred. NO. 1e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7

Db 441 CTNTSNP 447

RESULT 33

A11279

aconitate hydratases homolog citB [imported] - Listeria monocytogenes (strain EGD-e)

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 31-Dec-2004

C:Accession: A11279

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kretz, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluteter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlard,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11279
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-900 <GLA>
A:Cross-references: UNIPROT:Q8Y6P3; UNIPARC:UPI0000055009; GB:NC_003210; PIDN:CAC99719.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: citB
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 900;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 443 CTNTSNP 449

RESULT 34
A11642
aconitate hydratases homolog citB [imported] - *Listeria innocua* (strain Clip11262)
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 31-Dec-2004
C:Accession: A11642
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Faihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11642
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-900 <GLA>
A:Cross-references: UNIPROT:Q92B68; UNIPARC:UPI00000CC61P; GB:AL592022; PIDN:CAC96913.1;
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: citB
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 900;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 443 CTNTSNP 449

RESULT 35
F89910
aconitate hydratase [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 31-Dec-2004
C:Accession: F89910
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:2131952; PMID:11418146
A:Accession: F89910
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-901 <KUR>
A:Cross-references: UNIPROT:Q99UC8; UNIPARC:UPI00000549F4; GB:BA000018; PID:gl3701147; P93
A:Experimental source: strain N315
C:Genetics:
A:Gene: citB

C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 901;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 7
 ||||| ||
 Db 443 CTNTSNP 449

RESULT 36

A87704
 aconitate hydratase 1 [imported] - Caulobacter crescentus
 C;Species: Caulobacter crescentus
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 31-Dec-2004
 C;Accession: A87704
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of Caulobacter crescentus.
 A;Reference number: A87249; MUID:21173698; PMID:11259647
 A;Accession: A87704
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-903 <STO>
 A;Cross-references: UNIPROT:Q9A299; UNIPARC:UPI00000C7B68; GB:AE005673; NID:gl13425425; F
 C;Genetics:
 A;Gene: CC3667
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 903;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7

Db 444 CTNTSNP 450
 ||||| ||
 RESULT 37
 G75362
 aconitate hydratase - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Dec-2004
 C;Accession: G75362
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: G75362
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-906 <WHI>
 A;Cross-references: UNIPROT:Q9RTN7; UNIPARC:UPI00000D3EC3; GB:AE002013; GB:AE000513; NID
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR1720
 A;Map position: 1
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 906;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7

Db 441 CTNTSNP 447
 ||||| ||

RESULT 38

T04820
 aconitate hydratase (EC 4.2.1.3) F10M23.310 - Arabidopsis thaliana
 N;Alternate names: protein F10M23.310
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 31-Dec-2004
 C;Accession: T04820
 R;Bevan, M.; Lechamy, A.; Chedford, F.; Krivitzky, M.; Kreis, M.; Hoheisel, J.; Mewes, H.
 submitted to the Protein Sequence Database, February 1999
 A;Reference number: Z15385
 A;Accession: T04820
 A;Molecule type: DNA
 A;Residues: 1-907 <BEV>
 A;Cross-references: UNIPROT:Q9SZ36; UNIPARC:UPI00000AA6C2; EMBL:AL035440
 A;Experimental source: cultivar Columbia; BAC clone F10M23
 C;Genetics:
 A;Map position: 4
 A;Introns: 46/1; 99/3; 130/3; 192/3; 246/3; 295/1; 363/3; 388/2; 413/3; 475/3; 498/2; 531
 A;Note: F10M23.310
 C;Function:
 A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also to
 A;Pathway: glyoxylate bypass
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
 C;Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur prot

Query Match 62.3%; Score 38; DB 2; Length 907;

Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7

Db 450 CTNTSNP 456
 ||||| ||

RESULT 39

G82824
 aconitase XF0290 [imported] - Xylella fastidiosa (strain 945c)
 C;Species: Xylella fastidiosa
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 31-Dec-2004
 C;Accession: G82824
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; MUID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: G82824
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-908 <SIM>
 A;Cross-references: UNIPROT:Q9PGK8; UNIPARC:UPI00000C236F; GB:AE003883; GB:AE003849; NID:
 A;Experimental source: strain 945c
 R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; AJ
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Sa, R.G.; Santelli, R.V.; Sawasak
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A;Reference number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: XF0290
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 908;

Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
 |||||
 Db 443 CTNTSNP 449

RESULT 40
 G69599
 aconitate hydratase (EC 4.2.1.3) citB - Bacillus subtilis
 N/Alternate names: aconitase
 C/Species: Bacillus subtilis
 C/Accession: G69599; A27085
 R/Kunst, S.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emmerson, P.F.; Encian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A/Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A/Reference number: A69580; MUID:98044033; PMID:9384377
 A/Accession: G69599
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-909 <KUN>
 A/Cross-references: UNIPROT:P09339; UNIPARC:UPI0000060488; GB:Z99113; GB:AL009126; NID:9
 A/Experimental source: strain 168
 R/Dingman, D.W.; Sonenshein, A.L.
 J. Bacteriol. 169, 3062-3067, 1987
 A/Title: Purification of aconitase from Bacillus subtilis and correlation of its N-termi
 A/Reference number: A27085; MUID:87250270; PMID:3110133
 A/Accession: A27085
 A/Molecule type: DNA
 A/Residues: 1-28 'I', 30-42, 877-879 <DIN>
 A/Cross-references: UNIPARC:UPI0000176093
 A/Note: the authors translated the codon ATC for residue 29 as Tyr
 C/Genetics:
 A/Gene: citB
 C/Function:
 A/Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t
 A/Pathway: tricarboxylic acid cycle
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
 C/Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
 F:493-636/Domain: RNA binding #status predicted <RNA>
 F:135.136.188.219.221.316.792/Active site: Asp, His, Asp, His, Glu, Ser #status pre
 F:450.516.519/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 909;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
 |||||
 Db 450 CTNTSNP 456

RESULT 41
 B83451
 aconitate hydratase 1 PA1562 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2004
 C/Accession: B83451
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: B83451
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-910 <STO>
 A/Cross-references: UNIPROT:Q913F5; UNIPARC:UPI00000C53E0; GB:AE004584; GB:AE004091; NID:
 A/Experimental source: strain PAO1
 C/Genetics:
 A/Gene: acnA; PA1562
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 910;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
 |||||
 Db 454 CTNTSNP 460

RESULT 42
 F70873
 aconitate hydratase (EC 4.2.1.3) - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Dec-2004
 C/Accession: F70873
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230
 A/Accession: F70873
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-943 <COL>
 A/Cross-references: UNIPROT:O53166; UNIPARC:UPI00000D0F85; GB:AL021184; GB:AL123456; NID:
 A/Experimental source: strain H37RV
 C/Genetics:
 A/Gene: acn
 C/Function:
 A/Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t
 A/Pathway: tricarboxylic acid cycle
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
 C/Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
 F:522-665/Domain: RNA binding #status predicted <RNA>
 F:126.127.179.206.208.303.823/Active site: Asp, His, Asp, His, Glu, Ser #status pre
 F:479.545.548/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 943;
 Best Local Similarity 85.7%; Pred. No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
 |||||
 Db 479 CTNTSNP 485

RESULT 43
 G87135
 aconitate hydratase [imported] - Mycobacterium leprae
 C/Species: Mycobacterium leprae
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 31-Dec-2004
 C/Accession: G87135
 R/Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 sam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A/Title: Massive gene decay in the leprosy bacillus.
 A/Reference number: A86909; MUID:21128732; PMID:11234002

```
A:Accession: G87135
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-944 <STO>
A:Cross-references: UNIPROT:Q9CBL3; UNIPARC:UPI000000C6DFD; GB:AL450380; NID:gl3093525; F
C:Genetics:
A:Gene: acn
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 944;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7
    |||||
Db 480 CINTSNP 486

RESULT 44
A12205
hypothetical protein alr3200 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: A12205
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12205
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <KUR>
A:Cross-references: UNIPROT:Q8Y91; UNIPARC:UPI000000CB5DD; GB:BA000019; PIDN:BA074899.1;
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3200

Query Match 60.7%; Score 37; DB 2; Length 201;
Best Local Similarity 60.0%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 1; Indels 3; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
    |||||
Db 16 CQYRDNPYVE 25

RESULT 45
A96670
hypothetical protein F13011.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A96670
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizlar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96670
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <STO>
A:Cross-references: UNIPROT:Q9XIS2; UNIPARC:UPI000000A5918; GB:AE005173; NID:gs042406; PI
C:Genetics:
A:Gene: F13011.1
A:Map position: 1
```

```
Query Match 60.7%; Score 37; DB 2; Length 323;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNTDNPY 9
    |||||
Db 10 CTNTQNGRW 18

RESULT 46
I52603
MPS1 protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: I52603
R:Spilsbury, K.; O'Mara, M.A.; Wu, W.M.; Rowe, P.B.; Symonds, G.; Takayama, Y.
Blood 85, 1620-1629, 1995
A>Title: Isolation of a novel macrophage-specific gene by differential cDNA analysis.
A:Reference number: I52603; MUID:95195232; PMID:7888681
A:Accession: I52603
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-661 <RES>
A:Cross-references: UNIPROT:Q61889; UNIPARC:UPI0000002910D; GB:I20315; NID:g431419; PIDN:I

Query Match 60.7%; Score 37; DB 2; Length 661;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNTDNPY 9
    |||||
Db 349 CTNVDSNPF 357

RESULT 47
F64112
malate dehydrogenase (oxaloacetate-decarboxylating) homolog - Haemophilus influenzae (str
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 05-Oct-2004
C:Accession: F64112
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: F64112
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-756 <TIGR>
A:Cross-references: UNIPROT:P43837; UNIPARC:UPI00000512CB; GB:U32804; GB:I42023; NID:gl5;
C:Superfamily: malic enzyme with phosphate acetyl/butaryl transferase domain

Query Match 60.7%; Score 37; DB 2; Length 756;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NTDNPRYK 10
    |||||
Db 501 NEDNPRYE 508

RESULT 48
E88412
protein C44F1.5 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: E88412
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
```

A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E88412
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-805 <S>O>
A;Cross-references: UNIPROT:Q18628; UNIPARC:UPI00000769AA; GB:chr_III; PIDN:CAA88885.1;
C;Genetics:
A;Gene: C44F1.5
A;Map position: 3

Query Match 60.7%; Score 37; DB 2; Length 805;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NTQDPYK 10
| | | | |
| | | | |
Db 694 NTQDPYE 701

RESULT 49

T19936
hypothetical protein C44F1.5 - Caenorhabditis elegans (fragment)
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19936; T25216
R;Percy, C.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19199
A;Accession: T19936
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-805 <W>I>
A;Cross-references: UNIPROT:Q18628; UNIPARC:UPI00000769AA; EMBL:Z49067; PIDN:CAA88853.1;
A;Experimental source: clone C44F1
R;Percy, C.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19997
A;Accession: T25216
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-805 <W>I2>
A;Cross-references: UNIPARC:UPI00000769AA; EMBL:Z49072; PIDN:CAA88885.1; GSPDB:GN000021;
A;Experimental source: clone T24A11
C;Genetics:
A;Gene: CESP:C44F1.5
A;Map position: 3
A;Introns: 10/3; 60/3; 94/3; 227/2; 271/2; 314/2; 402/3; 434/3; 682/1

Query Match 60.7%; Score 37; DB 2; Length 805;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NTQDPYK 10
| | | | |
| | | | |
Db 694 NTQDPYE 701

RESULT 50

T37523
probable oxoprolinase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37523
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z21720
A;Accession: T37523
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1260 <CON>
A;Cross-references: UNIPROT:Q10093; UNIPARC:UPI000013A13B; EMBL:Z68166; PIDN:CAA92315.1;

A;Experimental source: strain 972h-; cosmid c11D3
C;Genetics:
A;Gene: SPDB:SPAC11D3.14c
A;Map position: 1
C;Superfamily: hypothetical protein YKL215c

Query Match 60.7%; Score 37; DB 2; Length 1260;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNTDNPY 9
| | | | |
| | | | |
Db 1220 CSNPDPFY 1228

Search completed: June 5, 2006, 12:54:04
Job time : 21.0959 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 87.6712 Seconds
(without alignments)
105.510 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CINTDNPRYK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Uniprot 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	543	1	HPSE HUMAN
2	49	80.3	545	1	Q9y251 homo sapien
3	43	70.5	567	2	Q3r2p1 xylella fas
4	43	70.5	608	2	Q3r890 xylella fas
5	43	70.5	861	2	Q3ql63 shewanella
6	43	70.5	862	2	Q33t54 shewanella
7	43	70.5	862	2	Q42UP2_PSEU2
8	43	70.5	862	2	Q88KF4_PSEPK
9	43	70.5	863	2	Q2X9A9_PSEPU
10	43	70.5	863	2	Q48JZ8_PSE14
11	43	70.5	863	2	Q2P7A3_XANOR
12	43	70.5	863	2	Q87P72_VIBPA
13	43	70.5	863	2	Q883R1_PSESM
14	43	70.5	863	2	Q8PND4_XANAC
15	43	70.5	864	2	Q2T374_BURTH
16	43	70.5	864	2	Q3BWH4_XANC5
17	43	70.5	864	2	Q454U1_9BURK
18	43	70.5	864	2	Q4BTS4_BURVI
19	43	70.5	864	2	Q4LQNS_9BURK
20	43	70.5	864	2	Q4URR6_XANC8
21	43	70.5	864	2	Q8VPS8_9BURK
22	43	70.5	864	2	Q39BA1_BURS3
23	43	70.5	864	2	Q3JH08_BURP1
24	43	70.5	864	2	Q3KFE7_PSEPF
25	43	70.5	864	2	Q4KFK1_PSEF5
26	43	70.5	864	2	Q5QZU0_IDILO
27	43	70.5	864	2	Q62A69_EURMA
28	43	70.5	864	2	Q63NU1_BURPS
29	43	70.5	864	2	Q8PBT6_XANCP
30	43	70.5	865	2	Q3F693_9BURK
31	43	70.5	865	2	Q3RQY1_RALME

32	43	70.5	865	2	Q8XTI8_RALSO
33	43	70.5	867	2	Q360W0_shewanella
34	43	70.5	867	2	Q36CY9_shewanella
35	43	70.5	867	2	Q3P203_9GAMM
36	43	70.5	867	2	Q4J5X3_AZCVI
37	43	70.5	867	2	Q8EJW3_shewanella
38	43	70.5	868	2	Q3CT85_pseudoalter
39	43	70.5	868	2	Q480I7_colwellia p
40	43	70.5	868	2	Q7NWD6_chromobacte
41	43	70.5	868	2	Q915E4_pseudomonas
42	43	70.5	868	2	Q9JT05_NEIMA
43	43	70.5	868	2	Q9K0X3_NEIMB
44	43	70.5	868	2	Q9KSC0_VIBCH
45	43	70.5	869	2	Q3IH91_PSEHT
46	43	70.5	869	2	Q3J7N8_RALEU
47	43	70.5	869	2	Q2KW75_BORAV
48	43	70.5	870	2	Q3NW39_SHEFR
49	43	70.5	871	2	Q2WXY8_9GAMM
50	43	70.5	871	2	Q2ZVF7_SHEPU
51	43	70.5	871	2	Q3QAM5_9GAMM
52	43	70.5	871	2	Q7ML96_VIBVU
53	43	70.5	871	2	Q8D975_VIBVU
54	43	70.5	872	2	Q34VQ9_9GAMM
55	43	70.5	874	2	Q36Q22_MARHY
56	43	70.5	877	2	Q8F8W5_ACIAD
57	43	70.5	881	2	Q470K8_RALEU
58	43	70.5	884	2	Q2ZAX7_9GAMM
59	43	70.5	885	2	Q3GR04_9GAMM
60	43	70.5	885	2	Q4FSP5_PSYAR
61	43	70.5	897	2	Q7VM95_BORPE
62	43	70.5	897	2	Q7WSQ5_BORPA
63	43	70.5	897	2	Q7WD91_BORBR
64	43	70.5	901	2	Q3RGR7_XYLFA
65	43	70.5	913	2	Q5H4H3_XANOR
66	43	70.5	939	2	Q7SD71_NEUCR
67	42	68.9	791	2	Q61KT7_CABER
68	41	67.2	82	2	Q66LT3_BPT5
69	41	67.2	87	2	Q6QGF5_BPT5
70	41	67.2	469	2	Q6FUB4_CANGA
71	41	67.2	558	2	Q333X5_SPAJD
72	41	67.2	574	2	Q333X6_SPAJD
73	41	67.2	574	2	Q333X7_9RODE
74	41	67.2	574	2	Q333X8_SPALAX GOLI
75	41	67.2	574	2	Q333X9_SPALAX GOLI
76	40	65.6	183	2	Q3CY93_STRAG
77	40	65.6	184	2	Q8XT50_RALSO
78	40	65.6	297	1	DAPA_LACIA
79	40	65.6	297	2	Q4A5E2_LACIA
80	40	65.6	338	2	Q2XSC0_MYTED
81	40	65.6	490	2	Q3DDM5_STRAG
82	40	65.6	551	2	Q3DC63_STRAG
83	40	65.6	881	2	Q3D064_STRAG
84	40	65.6	881	2	Q3D988_STRAG
85	40	65.6	881	2	Q3DJX7_STRAG
86	40	65.6	881	2	Q3DRB7_STRAG
87	40	65.6	881	2	Q3JZL9_STRA1
88	40	65.6	881	2	Q8DY23_STRA5
89	40	65.6	881	2	Q8E3P2_STRA3
90	40	65.6	960	1	SVV_BORBR
91	40	65.6	960	1	SVV_BORPA
92	40	65.6	960	1	SVV_BORPE
93	39	63.9	34	2	Q4T497_TETNG
94	39	63.9	145	2	Q6H520_ORYSA
95	39	63.9	235	2	Q4XEN9_PLACH
96	39	63.9	246	2	Q4X8D9_PLACH
97	39	63.9	259	2	Q97FY6_CLOAB
98	39	63.9	299	2	Q5FHS3_LACAC
99	39	63.9	536	1	HPSE_RAT
100	39	63.9	576	1	GLT12_MOUSE
101	39	63.9	576	2	Q60GT3_MOUSE
102	39	63.9	895	2	Q9Y040_PACLE
103	39	63.9	927	2	Q3WBW4_9ACTO
104	39	63.9	927	2	Q2JDF3_9ACTO

Q8xti8	ralstonia s
Q360w0	shewanella
Q36cy9	shewanella
Q3p203	shewanella
Q4j5x3	azotobacter
Q8ejw3	shewanella
Q3ct85	pseudoalter
Q480i7	colwellia p
Q7nwd6	chromobacte
Q915e4	pseudomonas
Q9jt05	neisseria m
Q9k0x3	neisseria m
Q9ksc0	vibrio chol
Q3ih91	pseudoalter
Q3j7n8	ralstonia e
Q2kw75	bordetella
Q3nw39	shewanella
Q2wxy8	shewanella
Q2zvf7	shewanella
Q3qam5	shewanella
Q7ml96	vibrio vuln
Q8d975	vibrio vuln
Q34vg9	alkalilimni
Q36q22	marinobacte
Q34vg9	alkalilimni
Q6f8w5	acinetobact
Q470k8	ralstonia e
Q2zax7	shewanella
Q3gr04	psychrobact
Q4fsp5	psychrobact
Q7vm95	bordetella
Q7wsq5	bordetella
Q7wd91	bordetella
Q3rgr7	xylella fas
Q5h4h3	xanthomonas
Q7sd71	neurospora
Q61kt7	caenorhabdi
Q66lt3	bacterioph
Q6qgt5	bacterioph
Q6fu84	candida gla
Q333x5	spalax juda
Q333x6	spalax juda
Q333x7	spalax carm
Q333x8	spalax goli
Q333x9	spalax goli
Q3cy93	streptococc
Q8xt50	ralstonia s
Q9cf61	lactococcus
Q54a52	lactococcus
Q2xsc0	mytilus edu
Q3ddm5	streptococc
Q3dc63	streptococc
Q3d064	streptococc
Q3d988	streptococc
Q3djx7	streptococc
Q3drb7	streptococc
Q3jzl9	streptococc
Q8dy23	streptococc
Q8e3p2	streptococc
Q7wm6	bordetella
Q7w6p3	bordetella
Q7vwk6	bordetella
Q4t497	tetraodon n
Q6h520	oryza sativ
Q4xen9	plasmodium
Q4x8d9	plasmodium
Q97fy6	clostridium
Q5fhs3	lactobacill
Q7ltp1	rattus norv
Q8bt9	m polypepti
Q60gt3	m uop-n-ace
Q9y040	pacificastac
Q3wbw4	frankia sp.
Q2jdf3	frankia sp.

689	34	55.7	121	2	Q9NM9_PLAFA	Q9nm9_plasmodium	762	34	55.7	411	2	Q40N13_DESAC	Q40n13_desulfuromo
690	34	55.7	132	2	Q8TUB5_METAC	Q8tub5_methanosarc	763	34	55.7	423	2	Q3SQ87_NITWN	Q3sq87_nitrobacter
691	34	55.7	136	2	Q9LZH8_ARATH	Q9lzh8_arabidopsis	764	34	55.7	430	2	Q9VW98_DROME	Q9vw98_drosophila
692	34	55.7	139	2	Q2RRA9_RHORA	Q2rra9_rhodospirill	765	34	55.7	431	2	Q21650_CABEL	Q21650_caenorhabdi
693	34	55.7	134	2	Q6Z102_ORYSA	Q6z102_oryza sativ	766	34	55.7	440	2	Q2V3J2_ARATH	Q2v3j2_arabidopsis
694	34	55.7	178	2	Q6S944_PLAFA	Q6s944_plasmodium	767	34	55.7	457	2	Q566C8_ORYSA	Q566c8_oryza sativ
695	34	55.7	178	2	Q4Y237_9HIV2	Q4y237_human immun	768	34	55.7	457	2	Q2S118_9GAMM	Q2s118_haella che
696	34	55.7	194	2	Q3YN09_BACTK	Q3yn09_bacillus th	769	34	55.7	465	2	Q9VGA7_DROME	Q9vga7_drosophila
697	34	55.7	199	2	Q4SD36_TETNG	Q4sd36_tetradodon n	770	34	55.7	469	1	Y1J1_CABEL	Y1j1_366 caenorhabdi
698	34	55.7	200	2	Q3VID3_CHLBA	Q3vid3_pelodictyon	771	34	55.7	469	2	Q6ACK3_LEIXX	Q6ack3_leifsonia x
699	34	55.7	200	2	Q4C3J1_CROWT	Q4c3j1_crocospaer	772	34	55.7	477	2	Q418U6_GIBZE	Q418u6_gibberella
700	34	55.7	216	2	Q46RG7_RALEU	Q46rg7_raistonia e	773	34	55.7	480	2	Q9LON4_STRCO	Q9lon4_streptomyce
701	34	55.7	224	2	Q2P835_XANOR	Q2p835_xanthomonas	774	34	55.7	485	1	SAH1_ARATH	Q23255_arabidopsis
702	34	55.7	224	2	Q5H5E0_XANOR	Q5h5e0_xanthomonas	775	34	55.7	485	1	SAH2_ARATH	Q91263_arabidopsis
703	34	55.7	240	2	Q9JVI1_STRCO	Q9jvi1_streptomyce	776	34	55.7	485	1	SAH1_MESCR	P93253_mesembryant
704	34	55.7	243	2	Q836J1_ENTPA	Q836j1_enterococcu	777	34	55.7	485	2	Q5D6C3_ARATH	Q5d6c3_arabidopsis
705	34	55.7	256	2	Q3CEA5_9CLOT	Q3cea5_alkaliphilu	778	34	55.7	485	2	Q5D6C4_ARATH	Q5d6c4_arabidopsis
706	34	55.7	260	2	Q3F2H0_9BURL	Q3f2h0_burkholderi	779	34	55.7	485	2	Q5D6C5_ARATH	Q5d6c5_arabidopsis
707	34	55.7	260	2	Q450T1_9BURL	Q450t1_burkholderi	780	34	55.7	488	2	Q4RJVO_TETNG	Q4rjvo_tetradodon n
708	34	55.7	260	2	Q4B8Z2_BURVI	Q4b8z2_burkholderi	781	34	55.7	489	2	Q7PNT2_ANOGA	Q7pnt2_anophelies g
709	34	55.7	260	2	Q4C3I9_CROWT	Q4c3i9_crocospaer	782	34	55.7	512	2	Q5YVS1_NOCFA	Q5yvs1_nocardia fa
710	34	55.7	260	2	Q4LNA7_9BURL	Q4lna7_burkholderi	783	34	55.7	517	1	MB13_YEAST	Q92zw7_saccharomyc
711	34	55.7	260	2	Q3PF58_BURS3	Q3pf58_burkholderi	784	34	55.7	527	2	Q6BNE9_DEBHA	Q6bne9_debaryomyc
712	34	55.7	262	2	Q62JD9_BURMA	Q62jd9_burkholderi	785	34	55.7	533	2	Q4RPR4_TETNG	Q4rpr4_tetradodon n
713	34	55.7	262	2	Q63T27_BURPS	Q63t27_burkholderi	786	34	55.7	533	2	Q7ZWU9_XENLA	Q7zwu9_xenopus lae
714	34	55.7	263	2	Q86276_ROT99	Q86276_bovine tota	787	34	55.7	534	2	Q6GPB8_XENLA	Q6gpb8_xenopus lae
715	34	55.7	266	2	Q37JN9_RHOPA	Q37jn9_bovine tota	788	34	55.7	534	2	Q6P886_XENTR	Q6p886_xenopus tro
716	34	55.7	271	2	Q8LWZ9_9PHAE	Q8lwz9_laminaria d	789	34	55.7	545	2	Q57VH9_9TRYP	Q57vh9_trypanosoma
717	34	55.7	271	2	Q36291_BIJOU	Q36291_bijou bridg	790	34	55.7	563	2	Q51D85_ENTHI	Q51d85_entamoeba h
718	34	55.7	274	2	Q6AHQ6_CABEL	Q6ahq6_caenorhabdi	791	34	55.7	589	2	Q43KQ2_SOLUS	Q43kq2_solibacter
719	34	55.7	274	2	Q4ZCK8_9CAUD	Q4zck8_bacterioph	792	34	55.7	591	2	Q62547_LOLPE	Q62547_loligo peal
720	34	55.7	274	2	Q4ZCS6_9CAUD	Q4zcs6_bacterioph	793	34	55.7	599	1	APAM_HUMAN	P43652_homo sapien
721	34	55.7	274	2	Q4ZD02_9CAUD	Q4zd02_bacterioph	794	34	55.7	599	2	Q32ME3_HUMAN	Q32mr3_homo sapien
722	34	55.7	274	2	Q8SDP2_9CAUD	Q8sdp2_staphylococ	795	34	55.7	602	2	Q9L112_STRCO	Q9l112_streptomyce
723	34	55.7	274	2	Q5HIY5_STAAC	Q5hiy5_staphylococ	796	34	55.7	603	2	Q7N4M8_PHOLL	Q7n4m8_photorhabdu
724	34	55.7	274	2	Q6GAK1_STAAS	Q6gak1_staphylococ	797	34	55.7	632	2	Q6F1M5_MESFL	Q6f1m5_mesoplasma
725	34	55.7	274	2	Q6GGF5_STAAR	Q6ggf5_staphylococ	798	34	55.7	643	1	Q1QRI1_RAT	Q9atf1_rattus norv
726	34	55.7	274	2	Q8NWK9_STAAM	Q8nwk9_staphylococ	799	34	55.7	651	2	Q7S668_NEUCR	Q7s668_neurospora
727	34	55.7	275	1	APAH_HAEN	P44751_haemophilus	800	34	55.7	656	2	Q7S668_NEUCR	Q7s668_neurospora
728	34	55.7	279	2	Q81EA3_BACCR	Q81ea3_bacillus ce	801	34	55.7	661	2	Q7FAF6_ORYSA	Q7faf6_oryza sativ
729	34	55.7	279	2	Q8EK17_SHEON	Q8ek17_shewanella	802	34	55.7	661	2	Q9PYR3_GVXN	Q9pyr3_xestia g-ni
730	34	55.7	281	2	Q2SWY3_BURTH	Q2swy3_burkholderi	803	34	55.7	665	2	Q41IW3_GIBZE	Q41iw3_gibberella
731	34	55.7	281	2	Q3JRA4_BURPL	Q3jra4_burkholderi	804	34	55.7	667	2	Q440G2_AERHY	Q440g2_aeromonas h
732	34	55.7	281	2	Q8AY27_YENLA	Q8ay27_xenopus lae	805	34	55.7	676	2	Q4UH19_THEAN	Q4uh19_thelateria a
733	34	55.7	289	2	Q8GGW7_LACLC	Q8ggw7_lactococcus	806	34	55.7	679	2	Q26820_9TRYP	Q26820_trypanosoma
734	34	55.7	289	2	Q9CJ12_LACLA	Q9cj12_lactococcus	807	34	55.7	679	2	Q389A6_9TRYP	Q389a6_trypanosoma
735	34	55.7	292	2	Q9GTL4_DROSI	Q9gtl4_drosophila	808	34	55.7	682	2	Q9STF4_ARATH	Q9stf4_arabidopsis
736	34	55.7	293	1	DAPA_STRMU	Q8dues_streptococc	809	34	55.7	712	2	Q7X8B0_ORYSA	Q7x8b0_oryza sativ
737	34	55.7	294	2	Q3KAT2_PSEPF	Q3kat2_pseudomonas	810	34	55.7	724	2	Q50QN7_ENTHI	Q50qn7_entamoeba h
738	34	55.7	305	2	Q2NX12_SODGL	Q2nx12_sodalis glo	811	34	55.7	729	2	Q5R112_IDILO	Q5r112_idiomarina
739	34	55.7	307	2	Q3NMV2_9GAMM	Q3nmv2_shewanella	812	34	55.7	733	2	Q6P7G8_XENLA	Q6p7g8_xenopus lae
740	34	55.7	308	2	Q4UQH2_XANC8	Q4uqh2_xanthomonas	813	34	55.7	751	1	SYV_NANEQ	Q74nf3_nanoarchaeu
741	34	55.7	308	2	Q8PCX8_XANCP	Q8pcx8_xanthomonas	814	34	55.7	767	2	Q623C4_CAEBR	Q623c4_caenorhabdi
742	34	55.7	311	2	Q75GK2_ORYSA	Q75gk2_oryza sativ	815	34	55.7	770	1	YRN9_CAEEL	Q09609_caenorhabdi
743	34	55.7	312	2	Q3VP27_9CHLB	Q3vp27_pelodictyon	816	34	55.7	777	2	Q2SJR4_9GAMM	Q2sjr4_haella che
744	34	55.7	320	2	Q7P5H5_FUSNV	Q7p5h5_fusobacteri	817	34	55.7	819	2	Q4WJ25_ASPFU	Q4wj25_aspergillus
745	34	55.7	320	2	Q8RHX5_FUSNV	Q8rhx5_fusobacteri	818	34	55.7	825	2	Q4RT16_TETNG	Q4rt16_tetradodon n
746	34	55.7	321	1	YKVZ_BACSU	Q31690_bacillus su	819	34	55.7	825	2	Q3CW15_ALTAT	Q3cw15_pseudoalter
747	34	55.7	327	2	Q85MB1_9FUNG	Q85mb1_monoblephar	820	34	55.7	912	2	Q9BDV6_ERIEU	Q9bdv6_erinaceus e
748	34	55.7	341	1	YSX3_CABEL	Q10022_caenorhabdi	821	34	55.7	914	2	Q9SD7Q_SOLTU	Q9sd7q_solanum tub
749	34	55.7	342	2	Q6AHQ7_CABEL	Q6ahq7_caenorhabdi	822	34	55.7	928	2	Q40PR3_DESAC	Q40pr3_desulfuromo
750	34	55.7	354	2	Q4HU81_GIBZE	Q4hu81_gibberella	823	34	55.7	932	2	Q2QVE3_ORYSA	Q2qvfe3_oryza sativ
751	34	55.7	361	2	Q5CIE4_CRYHO	Q5cie4_cryptospori	824	34	55.7	937	2	Q3AGF5_SOLTU	Q3agf5_solanum tub
752	34	55.7	362	2	Q8IOE1_DROME	Q8ioe1_drosophila	825	34	55.7	938	2	Q9SM52_SOLAC	Q9sm52_solanum aca
753	34	55.7	363	2	Q51AH0_ENTHI	Q51ah0_entamoeba h	826	34	55.7	941	2	Q4XY0_DICDI	Q4xy0_dicyostell
754	34	55.7	363	2	Q7MB10_PHOLL	Q7mb10_photorhabdu	827	34	55.7	947	2	Q8TSQ4_MANSE	Q8tsq4_manduca sex
755	34	55.7	364	2	Q2V3J3_ARATH	Q2v3j3_arabidopsis	828	34	55.7	952	2	Q6C6Q8_YARLI	Q6c6q8_yarrowia li
756	34	55.7	373	2	Q9FLV8_ARATH	Q9flv8_arabidopsis	829	34	55.7	993	2	Q4UFW1_THEAN	Q4ufw1_thelateria a
757	34	55.7	374	2	Q36758_YEAST	Q36758_saccharomyc	830	34	55.7	1041	2	Q9C2U8_GIBFU	Q9c2u8_gibberella
758	34	55.7	382	2	Q73XF7_MYCPA	Q73xf7_mycobacteri	831	34	55.7	1048	2	Q05925_NEIME	Q05925_neisseria m
759	34	55.7	385	2	Q6U7X9_CRIPE	Q6u7x9_crinipellis	832	34	55.7	1067	2	Q4P6A3_USTMA	Q4p6a3_ustilago ma
760	34	55.7	400	2	Q753U2_ASHGO	Q753u2_ashbya gos	833	34	55.7	1074	2	Q94046_CABEL	Q94046_caenorhabdi
761	34	55.7	402	2	Q4P4M6_USTMA	Q4p4m6_ustilago ma	834	34	55.7				

835	34	55.7	1076	2	Q9U370_CABEL	Q9U370	caenorhabdi	908	185	1	NUSG_HAEIN	P43916	haemophilus
836	34	55.7	1130	2	Q50WS9_ENTHI	Q50ws9	entamoeba h	909	185	2	Q3AW29_SYNS9	Q3aw29	synechococc
837	34	55.7	1159	2	Q4XU56_PLACH	Q4xus6	plasmodium	910	185	2	Q4QMK2_HAE18	Q4qmk2	haemophilus
838	34	55.7	1362	2	Q55CB6_DICDI	Q55cb6	dictyosteli	911	187	2	Q67282_AQUAE	Q67282	aquifex aeo
839	34	55.7	1437	1	VGLM_BUNGE	P12430	bunyavirus	912	188	2	Q7V4W1_PROMM	Q7v4w1	prochloroco
840	34	55.7	1588	2	Q4QAN5_LEIMA	Q4qan5	leishmania	913	190	2	Q7UE69_RHOBA	Q7ue69	rhodospirell
841	34	55.7	1647	2	Q54F23_DICDI	Q54f23	dictyosteli	914	191	2	Q3AM22_SYNSC	Q3am22	synechococc
842	34	55.7	1690	2	Q60YTA_CABER	Q60yta	caenorhabdi	915	193	2	Q4FAC8_9SALA	Q4fac8	eurycea gut
843	34	55.7	1901	2	Q9DHH8_YLDV	Q9dhh8	yaba-like d	916	199	2	Q6CUI4_KLULA	Q6cui4	kluyveromyc
844	34	55.7	2135	2	Q61077_PLAFA	Q61077	plasmodium	917	206	2	Q7V1M6_PROMP	Q7v1m6	prochloroco
845	34	55.7	2539	2	Q5F1P8_LACAC	Q5fip8	lactobacill	918	211	2	Q4G7D4_9TRYP	Q4g7d4	trypanosoma
846	34	55.7	3133	2	Q6BS09_DEBHA	Q6bs09	debaromyce	919	211	2	Q8Q209_9HIV1	Q8q209	human immun
847	33.5	54.9	105	2	Q9Q6M9_9DELA	Q9q6m9	human t-lym	920	212	2	Q8GEP8_9BACT	Q8gep8	uncultured
848	33.5	54.9	255	2	Q9D5S6_MOUSE	Q9d5s6	mus musculus	921	216	2	Q3HY53_9BACT	Q3hy53	uncultured
849	33.5	54.9	437	2	Q7RM38_PLAYO	Q7rm38	plasmodium	922	216	2	Q49CMA_9HIV1	Q49cm4	human immun
850	33.5	54.9	470	2	Q3TD11_MOUSE	Q3td11	mus musculus	923	224	1	VV_MUMPE1	P60167	mumps virus
851	33.5	54.9	504	2	Q3BDM1_PONPY	Q3bdm1	pongo pygma	924	224	1	VV_MUMPE	P30927	mumps virus
852	33.5	54.9	680	1	Y964_METUA	Q58374	methanococc	925	224	1	VV_MUMPM	P30928	mumps virus
853	33.5	54.9	702	2	Q8SY79_DROME	Q8sy79	drosophila	926	224	1	VV_MUMPS	P33483	mumps virus
854	33.5	54.9	702	2	Q9VKV6_DROME	Q9kvk6	drosophila	927	224	2	Q6B4U2_9PARA	Q6b4u2	mumps virus
855	33.5	54.9	873	2	Q4CGH7_CLOTM	Q4cgh7	clostridium	928	224	2	Q6BC98_9PARA	Q6bc98	mumps virus
856	33.5	54.9	882	2	Q8MTD1_NASBA	Q8mtd1	mastigamoeb	929	224	2	Q771S6_MUMPJ	Q771s6	mumps virus
857	33.5	54.9	1002	2	Q74674_PNEJI	Q74674	pneumocysti	930	224	2	Q833W0_9PARA	Q833w0	mumps virus
858	33.5	54.9	1162	2	Q7P9W1_RICSI	Q7p9w1	rickettsia	931	224	2	Q83620_9PARA	Q83620	mumps virus
859	33.5	54.9	1162	2	Q9DGA5_RICCN	Q9dga5	rickettsia	932	224	2	Q8QV69_MUMPJ	Q8qv69	mumps virus
860	33.5	54.9	1169	1	Y785_RICPR	O05975	rickettsia	933	224	2	Q91OG9_9PARA	Q91og9	mumps virus
861	33.5	54.9	1169	2	Q68VW3_RICTY	Q68vw3	rickettsia	934	224	2	Q9DQA4_9PARA	Q9dqa4	mumps virus
862	33.5	54.9	1528	1	ZFYI16_MOUSE	Q8ou44	mus musculus	935	224	2	Q9U4L5_9PARA	Q9u4l5	mumps virus
863	33	54.1	59	2	Q1JUB7_BACCR	Q1jub7	bacillus ce	936	225	1	ACE1_YEAST	P15315	saccharomyc
864	33	54.1	59	2	Q81W13_BACAN	Q81w13	bacillus an	937	228	2	Q6U955_9CAUD	Q6u955	aeromonas p
865	33	54.1	69	2	Q7MES6_VIBVY	Q7me56	vibrio vuln	938	229	2	Q8SML5_9CHLO	Q8sml5	dunaliella
866	33	54.1	86	1	SCR111_ARATH	P82630	arabidopsis	939	230	2	Q5E8Q9_VIBF1	Q5e8q9	vibrio fisc
867	33	54.1	103	2	Q9KAY6_BACHD	Q9kay6	bacillus ha	940	232	2	Q7NM03_GLOVI	Q7nm03	gloeobacter
868	33	54.1	109	2	Q4UUD7_RICFE	Q4uud7	rickettsia	941	238	2	Q54ED0_DICDI	Q54ed0	dictyosteli
869	33	54.1	126	2	Q8T4Y7_PLAFA	Q8t4y7	plasmodium	942	243	2	Q54ED0_DICDI	Q54ed0	dictyosteli
870	33	54.1	127	2	Q5DBB4_SCHJA	Q5dbb4	schistosoma	943	245	1	YAP5_YEAST	Q4ud44	theileria a
871	33	54.1	132	2	Q4DI06_TRYCR	Q4di06	trypanosoma	944	246	2	Q4XJ25_PLACH	Q4xj25	plasmodium
872	33	54.1	132	2	Q4DI06_TRYCR	Q4di06	trypanosoma	945	249	2	Q6AL84_DESPS	Q6al84	desulfotale
873	33	54.1	136	2	Q8TS09_PLAFA	Q8ts09	plasmodium	946	251	2	Q3LTN9_9GEMI	Q3ltn9	cabbage lea
874	33	54.1	140	2	Q5AJG7_CANAL	Q5ajg7	canidia alb	947	254	2	Q7PPX0_ANOGA	P10580	zebra mays (m
875	33	54.1	140	2	Q6Q3Q4_9COCO	Q6q3q4	infectious	948	256	1	YMS4_MAIZE	Q9mj33	zea mays (m
876	33	54.1	140	2	Q6Q3Q5_9COCO	Q6q3q5	infectious	949	256	2	Q9MJC3_MAIZE	Q9mj33	zea mays (m
877	33	54.1	140	2	Q6Q3Q6_9COCO	Q6q3q6	infectious	950	266	2	Q4I2A4_GIBZE	Q4i2a4	gibberella
878	33	54.1	140	2	Q6Q3Q8_9COCO	Q6q3q8	infectious	951	269	2	Q5WTM9_LEGPL	Q5wtm9	legionella
879	33	54.1	140	2	Q6Q3R1_9COCO	Q6q3r1	infectious	952	275	2	Q3VSP5_PROAE	Q3vsp5	prothetococh
880	33	54.1	140	2	Q6TN32_9COCO	Q6tn32	infectious	953	277	2	Q7P2Z0_FUSNV	Q7p2z0	fusobacteri
881	33	54.1	140	2	Q6TN33_9COCO	Q6tn33	infectious	954	288	2	Q6C1C4_YARLI	Q6c1c4	yarrowia li
882	33	54.1	140	2	Q6TN35_9COCO	Q6tn35	infectious	955	289	2	Q5NHB0_FRATT	Q5nhb0	franciella
883	33	54.1	140	2	Q6TN36_9COCO	Q6tn36	infectious	956	290	2	Q5DDM7_SCHJA	Q5ddm7	schistosoma
884	33	54.1	140	2	Q6TN37_9COCO	Q6tn37	infectious	957	290	2	Q7MJF5_VIBVY	Q7mjf5	vibrio vuln
885	33	54.1	140	2	Q6TN38_9COCO	Q6tn38	infectious	958	290	2	Q82LN9_STRAW	Q82ln9	streptomyce
886	33	54.1	140	2	Q6TN39_9COCO	Q6tn39	infectious	959	290	2	Q9S1Q5_STRCO	Q9s1q5	streptomyce
887	33	54.1	144	2	Q6TN39_9COCO	Q6tn39	infectious	960	290	2	Q8K2N4_MOUSE	Q8k2n4	mus musculu
888	33	54.1	145	2	Q9XWT0_CABEL	Q9xwt0	caenorhabdi	961	295	2	Q9H6Q3_HUMAN	Q9h6q3	homo sapien
889	33	54.1	145	2	Q9PT57_ANOGA	Q9pt57	anopheles g	962	300	2	Q8IEF4_PLAF7	Q8ief4	plasmodium
890	33	54.1	146	1	ANGI_COLGU	Q861y5	colobus gue	963	300	2	Q5Z9F9_ORISA	Q5z9f9	oryza sativ
891	33	54.1	149	2	Q7QT16_GIALA	Q7qtn63	macaca maca	964	300	2	Q8A974_BACTN	Q8a974	bacteroides
892	33	54.1	151	2	Q4HM34_CAMILA	Q4hm34	campylobact	965	300	2	Q8RG04_FUSNN	Q8rg04	fusobacteri
893	33	54.1	153	2	Q3VRN2_9SPHN	Q3vbn2	sphingopyxi	966	302	2	Q6BPK4_DBBHA	Q6bpk4	debaryomyce
894	33	54.1	154	2	Q3CAS9_9CLOT	Q3cas9	alkaliphilu	967	310	2	Q5BDW1_EMENI	Q5bdw1	aspergillu
895	33	54.1	163	2	Q4UIU9_9MEG	Q4uiu9	pungitius p	968	313	2	Q65N94_BACLD	Q65n94	bacillus li
896	33	54.1	164	2	Q23462_ARATH	Q23462	arabidopsis	969	314	2	Q6MYT0_ASPFU	Q6myt0	aspergillu
897	33	54.1	165	2	Q8TBR1_DROME	Q8tbr1	drosophila	970	315	2	Q69BT6_EMENI	Q69bt6	emericeila
898	33	54.1	172	1	IPYR_SYNEL	Q8dhr2	synechococc	971	316	2	Q2U2Y3_ASPOR	Q2u2y3	aspergillu
899	33	54.1	175	2	Q54VS3_DICDI	Q54vs3	dictyosteli	972	316	2	Q2WI52_CLOSTR	Q2wi52	clostridium
900	33	54.1	175	2	Q9M724_ARATH	Q9m724	arabidopsis	973	318	2	Q9PTX7_LAMRE	Q9ptx7	lametra re
901	33	54.1	179	2	Q4LJX2_ARATH	Q4ljx2	arabidopsis	974	319	2	Q3CAQ8_9CLOT	Q3caq8	alkaliphilu
902	33	54.1	180	2	Q4YV59_PLAAB	Q4yv59	plasmodium	975	321	2	Q9U5Q0_EPTBU	Q9u5q0	epatretus
903	33	54.1	180	2	Q7RRAS_PLAYO	Q7rras	plasmodium	976	326	2	Q5E8X8_VIBF1	Q5e8x8	vibrio fisc
904	33	54.1	181	1	COTE_BACSU	P14016	bacillus su	977	330	1	YBFG_ECOLI	P37749	escherichia
905	33	54.1	183	2	Q4I4M8_GIBZE	Q4i4m8	gibberella	978	330	2	Q380N7_ANOGA	Q380n7	anopheles g
906	33	54.1	184	2	Q38ZK2_LACSS	Q38zk2	lactobacill	979	331	2	Q3NEA3_HUMAN	Q3nea3	homo sapien
907	33	54.1	184	2	Q7U9H8_SYNPX	Q7u9h8	synechococc	980	331	2	Q9U2B4_CABEL	Q9u2b4	caenorhabdi

881 33 54.1 331 2 Q483Q7_COLP3 Q483q7 colwellia p
 882 33 54.1 332 2 Q3QEL7_9GAMM Q3qel7 shewanella
 883 33 54.1 333 2 Q3IF68_PSEHT Q3if68 pseudocalter
 884 33 54.1 334 1 ADD1_VIBPA Q87tf3 vibrio para
 885 33 54.1 334 1 ADD1_VIBPA Q9kn17 vibrio chol
 886 33 54.1 336 2 Q977B2_SULTO Q877b2 sulfolobus
 887 33 54.1 344 2 Q6BWE5_DEBHA Q8bwe5 debaryomyce
 888 33 54.1 345 2 Q9CM15_PASMU Q9cm15 pasteurella
 889 33 54.1 347 2 Q49YJ9_STAS1 Q49y9 staphylococ
 890 33 54.1 349 2 Q7JNV6_CAEEL Q7jnv6 caenorhabdi
 891 33 54.1 349 2 Q17377_CABEL Q17377 caenorhabdi
 892 33 54.1 349 2 Q34XY6_SGAMM Q34xy6 alkalilimni
 893 33 54.1 350 2 Q42928_SCHFO Q42928 schizosacch
 894 33 54.1 350 2 Q6LLR1_PHOPR Q6llr1 photobacter
 895 33 54.1 352 2 Q44BE2_SOLID Q44be2 solibacter
 896 33 54.1 352 2 Q6C4T3_YARLI Q6c4t3 yarrowia li
 897 33 54.1 357 2 Q41LM9_METHU Q41lm9 methanococ
 898 33 54.1 357 2 Q7JNV7_CAEEL Q7jnv7 caenorhabdi
 899 33 54.1 357 2 Q9DF34_BRARE Q9df34 brachydanio
 1000 33 54.1 359 1 DPO4_CLOPE Q8xk37 clostridium

ALIGNMENTS

RESULT 1
 HPSE HUMAN
 ID HPSE_HUMAN STANDARD; PRT; 543 AA.
 AC Q9Y251; Q53GE5; Q9UL39;
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
 DT 01-NOV-1999, sequence version 1.
 DT 07-FEB-2006, entry version 27.
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa
 DE subunit].
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRI, HPSE1, HSE1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=9935379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,
 RA Seddon A.P., Giorgio N.A., Bohlen P.;
 RT "Cloning and functional expression of a human heparanase gene.";
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
 [2]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.
 RC TISSUE=Embryonic fibroblast;
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
 RA Toyoshima M., Nakajima M.;
 RT "Human heparanase. Purification, characterization, cloning, and
 RT expression.";
 RL J. Biol. Chem. 274:24153-24160(1999).
 [3]
 RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
 RP PubMed=10395325; DOI=10.1038/10518;
 RX Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
 RA Spector L., Pecker I.;
 RT "Mammalian heparanase: gene cloning, expression and function in tumor
 RT progression and metastasis.";
 RL Nat. Med. 5:793-802(1999).
 [4]
 RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE
 RP OF 158-174; 263-272; 326-337; 433-436; 466-468 AND 478-483.
 RC TISSUE=Placenta;
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;
 RT "Cloning of mammalian heparanase, an important enzyme in tumor
 RT invasion and metastasis.";
 RL Nat. Med. 5:803-809(1999).
 [5]
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Placenta;
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RT "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage.";
 RL Glycobiology 10:467-475(2000).
 [6]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis.";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
 [7]
 RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Parekh R., Stubberfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpal) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435(2003).
 [8]
 RN NUCLEOTIDE SEQUENCE [MRNA].
 RP Pinhal M.A., Semedo P.;
 RA "Cloned heparanase from MCF-7 cells.";
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 [9]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP TISSUE=Small intestine;
 RC Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
 RA Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 [10]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP TISSUE=Pancras;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [11]
 RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RP PubMed=11123890; DOI=10.1021/bi002080p;
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Greedy J.E., Parish C.R.;

DR EMBL; AF281160; AAF87301.2; -; mRNA.
 DR InterPro; IPR005199; Glyco_hydro_79_N.
 DR Pfam; PF03662; Glyco_hydro_79n; 1.
 KW Calcium; Glycoprotein; Hydrolase; Lysozyme; Magnesium; Membrane;
 FT SIGNAL 1 37 By similarity.
 FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
 FT /FTid=PRO_0000042256.
 FT PROPEP 112 159 Linker peptide.
 FT /FTid=PRO_0000042257.
 FT CHAIN 160 545 Heparanase 50 kDa subunit (By similarity).
 FT /FTid=PRO_0000042258.
 FT REGION 160 164 Heparin/HS-binding (Potential).
 FT REGION 272 282 Heparin/HS-binding (Potential).
 FT ACT SITE 227 227 Proton donor (Potential).
 FT ACT SITE 345 345 Nucleophile (Potential).
 FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;

Query Match 80.3%; Score 49; DB 1; Length 545;
 Best Local Similarity 80.0%; Pred. No. 4;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
 |||||:||||
 DB 439 CTNTKHPYK 448

RESULT 3
 Q3R2P1_XYLFA
 ID Q3R2P1_XYLFA PRELIMINARY; PRT; 567 AA.
 AC Q3R2P1;
 DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
 DT 25-OCT-2005, sequence version 1.
 DT 07-FEB-2006, entry version 3.
 DE Aconitate hydratase (EC 4.2.1.3) (Fragment).
 GN ORFNames=XfasoDRAFT_1379;
 OS Xylella fastidiosa Ann-1.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=155920;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.,
 RA "Sequencing of the draft genome and assembly of Xylella fastidiosa
 RT Ann-1";
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RA "Annotation of the draft genome assembly of Xylella fastidiosa Ann-
 RT 1";
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.

CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; AAAM0300062; EAQ31487.1; -; Genomic DNA.
 DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR000573; Aconitase_C.
 DR InterPro; IPR001030; Aconitase_N.
 DR Pfam; PF00330; Aconitase; 1.
 DR Pfam; PF00694; Aconitase C; 1.
 DR PRINTS; PR00415; ACONITASE.
 DR ProDom; PD000511; Aconitase_N; 1.
 KW Lyase.
 FT NON TER 1
 SQ SEQUENCE 567 AA; 61337 MW; 8F9E4B7AFCE7017E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 567;
 Best Local Similarity 87.5%; Pred. No. 51;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
 |||||
 DB 104 CTNTSNP 111

RESULT 4
 Q3R890_XYLFA
 ID Q3R890_XYLFA PRELIMINARY; PRT; 608 AA.
 AC Q3R890;
 DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
 DT 25-OCT-2005, sequence version 1.
 DT 07-FEB-2006, entry version 3.
 DE Aconitate hydratase, N-terminal.
 GN ORFNames=XfasoDRAFT_3065;
 OS Xylella fastidiosa Ann-1.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=155920;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA "Sequencing of the draft genome and assembly of Xylella fastidiosa
 RT Ann-1";
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RA "Annotation of the draft genome assembly of Xylella fastidiosa Ann-
 RT 1";
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Ann-1;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
 RL -!- CAUTION: The sequence shown here is derived from an

```
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAAM03000011; EAO33483.1; -; Genomic_DNA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR GO; GO:0008152; P:metabolism; IEA.
CC DR InterPro; IPR001030; Aconitase_N.
CC DR Pfam; PF00330; Aconitase; 1.
CC DR PRINTS; PR00415; ACONITASE.
CC DR ProDom; PD000511; Aconitase_N; 1.
CC SQ SEQUENCE 608 AA; 64732 MW; FDIAB95AD0A45DD4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 608;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 438 CTNTSNPR 445

RESULT 5
Q3QL63_9GAMM PRELIMINARY; PRT; 861 AA.
AC Q3QL63;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
GN ORFNames=SanaDRAFT_1319;
OS Shewanella amazonensis SB2B.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=326297;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB2B;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Shewanella amazonensis
RT SB2B.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB2B;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Shewanella amazonensis
RT SB2B.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAIN01000001; EAA40876.1; -; Genomic_DNA.
CC DR GO; GO:0016836; F:hydro-lyase activity; IEA.
CC DR GO; GO:0005506; F:iron ion binding; IEA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR InterPro; IPR012708; 2met_isocit_dhvd.
CC DR InterPro; IPR012084; Aco_LysF.
CC DR InterPro; IPR000573; Aconitase_C.
CC DR Pfam; PF00330; Aconitase; 1.
CC DR Pfam; PF00694; Aconitase_C; 1.
CC DR PIRSF; PIRSF001417; LysF; 1.
```

```
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhvy; 1.
SQ SEQUENCE 861 AA; 94219 MW; 47750044C9B194A CRC64;

Query Match 70.5%; Score 43; DB 2; Length 861;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413

RESULT 6
Q33T54_9GAMM PRELIMINARY; PRT; 862 AA.
AC Q33T54;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
GN ORFNames=ShewDRAFT_0853;
OS Shewanella sp. PV-4.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=323850;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Shewanella sp. PV-4.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Shewanella sp. PV-4.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALS01000019; EAP02293.1; -; Genomic_DNA.
CC DR GO; GO:0016836; F:hydro-lyase activity; IEA.
CC DR GO; GO:0005506; F:iron ion binding; IEA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR GO; GO:0008152; P:metabolism; IEA.
CC SQ SEQUENCE 862 AA; 94331 MW; 131DD5B7AE952023 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413

RESULT 7
Q4ZUP2_PSEU2 PRELIMINARY; PRT; 862 AA.
AC Q4ZUP2;
DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-JUN-2005, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
```

```

GN OrderedLocusNames=Psy_2087;
OS Pseudomonas syringae pv. syringae (strain B728a).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043691; DOI=10.1073/pnas.0504930102;
RA Feil H., Feil W.S., Chain P., Larimer F., Dibartolo G., Copeland A.,
RA Lykidis A., Trong S., Nolan M., Goltzman E., Thiel J., Malfatti S.,
RA Loper J.E., Lapidus A., Detter J.C., Land M., Richardson P.M.,
RA Kyrpides N.C., Ivanova N., Lindov S.E.;
RT "Comparison of the complete genome sequences of Pseudomonas syringae
RT pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 102:11064-11069(2005).
RL
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
DR EMBL; CP000075; AAY37130.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 862 AA; 93915 MW; C89D1AD262127EBA CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 8
Q88KF4_PSEPK PRELIMINARY; PRT; 862 AA.
AC Q88KF4;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Aconitate hydratase, putative.
GN OrderedLocusNames=PP2336; ORFNames=PP_2336;
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22423020; PubMed=12534463;
RX DOI=10.1046/j.1462-2920.2002.00366.x;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapfel E.K., Scanlan D., Tran K.,
RA Moazzaz A., Utecherback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA Wedler H., Lauber J., Stjepandic D., Hohenseil J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."

```

```

RL Environ. Microbiol. 4:799-808(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC
DR EMBL; AE015451; AAN67949.1; -; Genomic DNA.
DR TIGR; PP2336; -;
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 862 AA; 93962 MW; BCB84EA2AE8624B6 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 9
Q2X9A9_PSEPU PRELIMINARY; PRT; 863 AA.
AC Q2X9A9;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, N-terminal.
GN ORFNames=PputDRAFT_0339;
OS Pseudomonas putida Fl.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=351746;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Fl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P., Richardson P.,
RT "Sequencing of the draft genome and assembly of Pseudomonas putida
RT Fl."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Fl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Pseudomonas putida Fl.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AALM01000071; EAP48361.1; -; Genomic DNA.
SQ SEQUENCE 863 AA; 94162 MW; 82F2F8549795839C CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;

```

```
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 407 CTNTSNR 414

RESULT 10
Q48JZ8_PSE14 PRELIMINARY; PRT; 863 AA.
AC Q48JZ8;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DT 07-FEB-2006, entry version 5.
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).
GN Name-acnd; OrderedLocNames=PSPH 2058;
OS Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=264730;
RN 1 CTNTDNR 8
RX PubMed=16159782; DOI=10.1128/JB.187.18.6488-6498.2005;
RA Joardar V., Lindeberg M., Jackson R., Selengut J., Dodson R.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA Gwinn Gligio M., Madupu R., Nelson W.C., Rosovitz M.J., Sullivan S.A.,
RA Crabtree J., Creasy T., Davidson T.M., Haft D.H., Zafar N., Zhou L.,
RA Halpin R., Holley T., Khouri H.M., Feldblyum T.V., White O.,
RA Fraser C.M., Chatterjee A.K., Cartinhour S., Schneider D.,
RA Mansfield J., Collmer A., Buell R.;
RT "Whole-genome sequence analysis of Pseudomonas syringae pv.
RT phaseolicola 1448A reveals divergence among pathogens in genes
RT involved in virulence and transposition.";
RL J. Bacteriol. 187:6488-6498(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: CP000058; AA233624.1; -; Genomic DNA.
DR GO: GO:0047456; F-2-methylisocitrate dehydratase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 863 AA; 94098 MW; 18D906940783C488 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 407 CTNTSNR 413

RESULT 11
Q2P7A3_XANOR PRELIMINARY; PRT; 863 AA.
AC Q2P7A3;
DT 07-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DT 07-FEB-2006, entry version 1.
DE Aconitase hydratase 1.
SQ SEQUENCE 863 AA; 94791 MW; ACP4454F1C74B97B CRC64;
```

```
Name=XO00819;
Xanthomonas oryzae pv. oryzae MAFF 311018.
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=342109;
RN 1 CTNTDNR 8
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MAFF 311018;
RA Ochiai H., Inoue Y., Takeya M., Sasaki A., Kaku H.;
RT "Genome sequence of Xanthomonas oryzae pv. oryzae suggests
RT contribution of large numbers of effector genes and insertion
RT sequences to its race diversity.";
RL Jpn. Agric. Res. Q. 39:275-287(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: AP008229; BAE67574.1; -; Genomic DNA.
DR EMBL; AP008229; BAE67574.1; -; Genomic DNA.
SQ SEQUENCE 863 AA; 93934 MW; 93D16B0BE4CAF67B CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 406 CTNTSNR 413

RESULT 12
Q87P72_VIBPA PRELIMINARY; PRT; 863 AA.
AC Q87P72;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 15.
DE Aconitase hydratase 1.
GN OrderedLocNames=VPI646;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN 1 CTNTDNR 8
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nishima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: BA000031; BACS9909.1; -; Genomic DNA.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 863 AA; 94791 MW; ACP4454F1C74B97B CRC64;
```


Query Match 70.5%; Score 43; DB 2; Length 863;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 8
 DB 410 CTNTSNP 417

RESULT 13
 Q883R1_PSESM PRELIMINARY; PRT; 863 AA.
 AC Q883R1;
 DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
 DT 01-JUN-2003, sequence version 1.
 DT 07-FEB-2006, entry version 13.
 DE Aconitase family protein.
 GN OrderedLocusNames=PSPTO2289; ORFNames=PSPTO_2289;
 OS Pseudomonas syringae pv. tomato.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=323;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=DC3000;
 RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
 RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
 RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
 RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
 RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhou S., Chatterjee A.K.,
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
 RA Bender C.L., White O., Fraser C.M., Collier A.;
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen
 RT Pseudomonas syringae pv. tomato DC3000.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186 (2003).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AS016853; AA055803.1; -; Genomic_DNA.
 DR TIGR; PSPT02289;
 DR BioCyc; PSYR223283:PSPTO2289-MONOMER; -; IEA.
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.
 DR GO; GO:0005506; F:iron ion binding; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR012708; 2met_isocit_dHVD.
 DR InterPro; IPR012084; Aco_LysF.
 DR InterPro; IPR000573; Aconitase_C.
 DR InterPro; IPR001030; Aconitase_N.
 DR Pfam; PF00694; Aconitase; 1.
 DR Pfam; PF00694; Aconitase; 1.
 DR PIRSF; PIRSF001417; LysF; 1.
 DR PRINTS; PD00415; ACONITASE.
 DR ProDom; PD000511; Aconitase_N; 1.
 DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
 DR Complete proteome.
 KW SEQUENCE 863 AA; 94052 MW; 6580FB5E2ACE13E0 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 8
 DB 406 CTNTSNP 413

RESULT 14

Q8PND4_XANAC
 ID Q8PND4_XANAC PRELIMINARY; PRT; 863 AA.
 AC Q8PND4;
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
 DT 01-OCT-2002, sequence version 1.
 DT 07-FEB-2006, entry version 13.
 DE Aconitase hydratase 1.
 GN Name=acnA;
 OS Xanthomonas axonopodis pv. citri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=92829;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A.,
 RA Almeida N.F. Jr., Alves L.M.C., do Amaral A.M., Bertolini M.C.,
 RA Camargo L.E.A., Camarotte G., Cannavan F., Cardozo J., Chambergo F.,
 RA Ciapina L.P., Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R.,
 RA El-Dorry H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Gruber A.,
 RA Ferro M.I.T., Formighieri E.F., Franco M.C., Greggio C.C., Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities.";
 RL Nature 417:459-463 (2002).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AE011742; AAM36011.1; -; Genomic DNA.
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.
 DR GO; GO:0005506; F:iron ion binding; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR012708; 2met_isocit_dHVD.
 DR InterPro; IPR012084; Aco_LysF.
 DR InterPro; IPR000573; Aconitase_C.
 DR InterPro; IPR001030; Aconitase_N.
 DR Pfam; PF00694; Aconitase; 1.
 DR Pfam; PF00694; Aconitase; 1.
 DR PIRSF; PIRSF001417; LysF; 1.
 DR PRINTS; PD00415; ACONITASE.
 DR ProDom; PD000511; Aconitase_N; 1.
 DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
 DR Complete proteome.
 KW SEQUENCE 863 AA; 93708 MW; 7504D1917BD80F4E CRC64;
 SQ
 Query Match 70.5%; Score 43; DB 2; Length 863;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 8
 DB 406 CTNTSNP 413

RESULT 15
 Q2T374_BURTH
 ID Q2T374_BURTH PRELIMINARY; PRT; 864 AA.
 AC Q2T374;
 DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
 DT 24-JAN-2006, sequence version 1.
 DT 07-FEB-2006, entry version 2.
 DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).

```
GN Name=acnD; ORFNames=BTH_I12187;
OS Burkholderia thailandensis E264;
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=271848;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=E264;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Palmer N., Dodson R.,
RA Hickey E.K., Gwinn M., Dougherty B., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van-Vugt R., Adams M.D., Gocayne J.D., Weidman J., Uterback T.,
RA Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C.,
RA Cotton M.D., Horst K., Tomb J.-F., Roberts K., Hatch B., Smith H.O.,
RA Venter J.C.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000085; ABC34711.1; -; Genomic DNA.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase.
SQ SEQUENCE 864 AA; 94223 MW; 4C12E4081417726 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 16
Q3BWH4_XANC5 PRELIMINARY; PRT; 864 AA.
AC Q3BWH4;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Aconitate hydratase (EC 4.2.1.3).
GN Name=acnA; OrderedLocusNames=XCV1158;
OS Xanthomonas campestris pv. vesicatoria (strain 85-10).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=316273;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16237009; DOI=10.1128/JB.187.21.7254-7266.2005;
RA Thiene F., Koebnik R., Bekel T., Berger C., Boch J., Buettner D.,
RA Chaldana C., Gaigalat L., Goemann A., Kay S., Kirchner O., Lanz C.,
RA Linke B., McHardy A.C., Meyer F., Mittenhuber G., Nies D.H.,
RA Niesbach-Kloeegen U., Patschewski T., Rueckert C., Rupp O.,
RA Schniesker S., Schuster S.C., Vorhoefer F.J., Weber E., Puehler A.,
RA Bonas U., Bartels D., Kaiser O.;
RT "Insights into genome plasticity and pathogenicity of the plant
RT pathogenic bacterium Xanthomonas campestris pv. vesicatoria revealed
RT by the complete genome sequence.";
RL J. Bacteriol. 187:7254-7266 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AM039952; CAJ22789.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 94066 MW; ADBAD8BA90C6BDE7 CRC64;

Name=acnD; ORFNames=BTH_I12187;
OS Burkholderia thailandensis E264;
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=271848;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=E264;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Palmer N., Dodson R.,
RA Hickey E.K., Gwinn M., Dougherty B., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van-Vugt R., Adams M.D., Gocayne J.D., Weidman J., Uterback T.,
RA Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C.,
RA Cotton M.D., Horst K., Tomb J.-F., Roberts K., Hatch B., Smith H.O.,
RA Venter J.C.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000085; ABC34711.1; -; Genomic DNA.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase.
SQ SEQUENCE 864 AA; 94223 MW; 4C12E4081417726 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 17
Q454U1_9BURK PRELIMINARY; PRT; 864 AA.
AC Q454U1;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=BcenDRAFT_3590;
OS Burkholderia cenocepacia AU 1054.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cenocepacia complex.
OX NCBI_TaxID=331271;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AU 1054;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AU 1054;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAHT01000012; EMML2040.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met.isocit.dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met.isocit.dhy; 1.
SQ SEQUENCE 864 AA; 93860 MW; 60CFA11E2A0133C9 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 18
Q4BTS4_BURVI
```

ID Q4BTS4_BURVI PRELIMINARY; PRT; 864 AA.
 AC Q4BTS4;
 DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
 DT 13-SEP-2005, sequence version 1.
 DT 07-FEB-2006, entry version 2.
 DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
 GN ORNames=Bcep1808DRAFT_7478;
 OS Burkholderia vietnamiensis G4.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
 OX NCBI_TaxID=269462;
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=G4;
 RC US DOE Joint Genome Institute (JGI-PGF);
 RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.,
 RA "Sequencing of the draft genome and assembly of Burkholderia
 RT vietnamiensis G4";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=G4;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RA "Annotation of the draft genome assembly of Burkholderia vietnamiensis
 RT G4";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=G4;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NonDerivs License
 CC -----
 DR EMBL; AAHE02000001; EAM33209.1; -; Genomic DNA.
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.
 DR GO; GO:0005506; F:iron ion binding; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR012708; 2met_isocit_dHYD.
 DR InterPro; IPR012084; Aco_LysF.
 DR InterPro; IPR01030; Aconitase_N.
 DR Pfam; PF00330; Aconitase; 1.
 DR Pfam; PF00694; Aconitase_C; 1.
 DR PIRSF; PIRSF001417; LysF; 1.
 DR PRINTS; PR00415; ACONITASE.
 DR ProDom; PD000511; Aconitase_N; 1.
 DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
 DR SEQUENCE 864 AA; 94035 MW; 93B5A360E67F0FB1 CRC64;
 SQ
 Query Match 70.5%; Score 43; DB 2; Length 864;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 8
 |||||
 Db 406 CTNTSNPR 413
 RESULT 19
 Q4LQNS_9BURK PRELIMINARY; PRT; 864 AA.
 ID Q4LQNS_9BURK
 AC Q4LQNS;
 DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.
 DT 02-AUG-2005, sequence version 1.

DT 07-FEB-2006, entry version 2.
 DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
 GN ORNames=Bcen2424DRAFT_3126;
 OS Burkholderia cenocepacia HI2424.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
 OX NCBI_TaxID=331272;
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=HI2424;
 RC US DOE Joint Genome Institute (JGI-PGF);
 RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pitluck S., Richardson P.;
 RA "Sequencing of the draft genome assembly of Burkholderia cenocepacia
 RT HI2424";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=HI2424;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RA "Annotation of the draft genome assembly of Burkholderia cenocepacia
 RT HI2424";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NonDerivs License
 CC -----
 DR EMBL; AAHL01000027; EAM18383.1; -; Genomic DNA.
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.
 DR GO; GO:0005506; F:iron ion binding; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR012708; 2met_isocit_dHYD.
 DR InterPro; IPR012084; Aco_LysF.
 DR InterPro; IPR01030; Aconitase_N.
 DR Pfam; PF00330; Aconitase; 1.
 DR Pfam; PF00694; Aconitase_C; 1.
 DR PIRSF; PIRSF001417; LysF; 1.
 DR PRINTS; PR00415; ACONITASE.
 DR ProDom; PD000511; Aconitase_N; 1.
 DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
 DR SEQUENCE 864 AA; 93860 MW; 60CFA11E2A0133C9 CRC64;
 SQ
 Query Match 70.5%; Score 43; Length 864;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTNTDNP 8
 |||||
 Db 406 CTNTSNPR 413
 RESULT 20
 Q4URR6_XANC8 PRELIMINARY; PRT; 864 AA.
 ID Q4URR6_XANC8
 AC Q4URR6;
 DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
 DT 05-JUL-2005, sequence version 1.
 DT 07-FEB-2006, entry version 5.
 DE Aconitate hydratase 1.
 GN OrderedLocusNames=XC_3213;
 OS Xanthomonas campestris pv. campestris (strain 8004).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=314565;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RX PubMed=15899963; DOI=10.1101/gr.3378705;
 RA Qian W., Jia Y., Ren S.-X., He Y.-Q., Feng J.-X., Lu L.-F., Sun Q.,

RA Ying G., Tang D.-J., Tang H., Wu W., Hao P., Wang L., Jiang B.-L.,
RA Zeng S., Gu W.-Y., Lu G., Rong L., Tian Y., Yao Z., Fu G., Chen B.,
RA Fang R., Qiang B., Chen Z., Zhao G.-P., Tang J.-L., He C.,
RT "Comparative and functional genomic analyses of the pathogenicity of
RL phytopathogen *Xanthomonas campestris* pv. *campestris*.";
CC Genome Res. 15:757-767(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000050; AAY50257.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 864 AA; 94059 MW; 8CC8BBBA4E1696BF CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 21
Q8VPS8_9BURK
ID Q8VPS8_9BURK PRELIMINARY; PRT; 864 AA.
AC Q8VPS8;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE AcnM.
GN Name=acnM;
OS Burkholderia sacchari.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=159450;
RN [1]
R1 NUCLEOTIDE SEQUENCE.
RP STRAIN=IPT101T;
RC MEDLINE=21633825; PubMed=11772636; DOI=10.1128/AEM.68.1.271-279.2002;
RA Bramer C.O., Silva L.F., Gomez J.G.C., Priefer H., Steinbuechel A.,
RT "Identification of the 2-methylcitrate pathway involved in the
RT catabolism of propionate in the Polyhydroxyalkanoate-Producing Strain
RT Burkholderia sacchari IPT101(T) and Analysis of a Mutant Accumulating
RT a Copolyester with Higher 3-Hydroxyvalerate Content.";
RL Appl. Environ. Microbiol. 68:271-279(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IPT101T;
RA Braemer C.O., Steinbuechel A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AY033092; AAK52341.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase N; 1. dhy; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 864 AA; 94376 MW; A058CED3BC95C48C CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 22
Q39BAL_BURS3
ID Q39BAL_BURS3 PRELIMINARY; PRT; 864 AA.
AC Q39BAL;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE Aconitate hydratase-like (EC 4.2.1.3).
GN OrderedLocusNames=Bcep18194_B0143;
OS Burkholderia sp. (strain 383) (Burkholderia cepacia (strain ATCC 17760
/ NCIB 9086 / R18194)).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=269483;
RN [1]
R1 NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Istrani S., Pitluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Kyrpides N., Lykidis A.,
RA Richardson P.;
RA "Complete sequence of chromosome 2 of Burkholderia sp. 383.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000152; ABB10260.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
KW Complete proteome; lyase.
SQ SEQUENCE 864 AA; 94002 MW; AD8FC6C259002FE4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 23
Q3JHQ8_BURP1
ID Q3JHQ8_BURP1 PRELIMINARY; PRT; 864 AA.
AC Q3JHQ8;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).

```
GN Name-acnd; OrderedLocusNames=BURPS1710b_A1738;
OS Burkholderia pseudomallei (strain 1710b).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=320372;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Woods D.E., Nierman W.C.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000125; ABA53629.1; -; Genomic_DNA.
DR TIGR; BURPS1710b_A1738; -;
DR GO; GO:0047456; F:2-methylisocitrate dehydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 93924 MW; F8D1F356678CB2B5 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413
|||||

RESULT 24
Q3KFE7_PSEPF
ID Q3KFE7_PSEPF PRELIMINARY; PRT; 864 AA.
AC Q3KFE7;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE Aconitase hydratase-like.
GN OrderedLocusNames=Pfl_1766;
OS Pseudomonas fluorescens (strain pfo-1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205922;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Saunders E.H., Schmutz J.,
RA Larriner F., Land M., Kyripiides N., Anderson I., Richardson P.;
RA "Complete sequence of Pseudomonas fluorescens pfo-1";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000094; ABA73509.1; -; Genomic_DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
```

```
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 864 AA; 94414 MW; 057777344276BD90 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413
|||||

RESULT 25
Q4KFK1_PSEF5
ID Q4KFK1_PSEF5 PRELIMINARY; PRT; 864 AA.
AC Q4KFK1;
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 02-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).
GN Name=acnd; OrderedLocusNames=PFL1863;
OS Pseudomonas fluorescens (strain Pf-5 / ATCC BAA-477).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=220664;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15980861; DOI=10.1038/nbt1110;
RA Paulsen I.T., Press C.M., Ravel J., Kobayashi D.Y., Myers G.S.A.,
RA Mavrodi D.V., DeBoy R.T., Seshadri R., Ren Q., Madupu R., Dodson R.J.,
RA Durkin A.S., Brinkac L.M., Daugherty S.C., Sullivan S.A.,
RA Rosovitz M.J., Gwinn M.L., Zhou L., Schneider D.J., Cartinhour S.W.,
RA Nelson W.C., Weidman J., Watkins K., Tran K., Khouri H., Pierson E.A.,
RA Pierson L.S. III, Thomashow L.S., Loper J.E.;
RT "Complete genome sequence of the plant commensal Pseudomonas
fluorescens Pf-5.";
RL Nat. Biotechnol. 23:873-878(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000076; AAY91151.1; -; Genomic_DNA.
DR GO; GO:0047456; F:2-methylisocitrate dehydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 94154 MW; C8F37A1B7BAB56D2 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413
|||||
```

```
RESULT 26
QS0ZUO_IDILO PRELIMINARY; PRT; 864 AA.
ID QS0ZUO;
AC QS0ZUO;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Aconitase A.
GN OrderedLocusNames=IL1425;
OS Idiomarina loihiensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Idiomarinaceae; Idiomarina.
OX NCBI_TaxID=135577;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=L2-TR / DSM 15497 / ATCC BAA-735;
RX PubMed=15596722; DOI=10.1073/pnas.0407638102;
RA Hou S., Saw J.H., Lee K.S., Freitas T.A., Bellisle C., Kawarabayashi Y.,
RA Donachie S.P., Fikina A., Galperin M.Y., Koonin E.V., Makarova K.S.,
RA Omelchenko M.V., Sorokin A., Wolf Y.I., Li Q.X., Keum Y.S.,
RA Campbell S., Denery J., Aizawa S., Shibata S., Malahoff A., Alam M.;
RA "Genome sequence of the deep-sea gamma-proteobacterium Idiomarina
RT loihiensis reveals amino acid fermentation as a source of carbon and
RT energy."
RL Proc. Natl. Acad. Sci. U.S.A. 101:18036-18041(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR ENBL; AE017340; AAV82265.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR00573; Aconitase C.
DR Pfam; PF00330; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome.
SQ SEQUENCE 864 AA; 95071 MW; E53C54A825E79B6B CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 406 CTNTSNP 413

RESULT 27
Q62A69_BURMA PRELIMINARY; PRT; 864 AA.
ID Q62A69_BURMA
AC Q62A69;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Aconitase hydratase (EC 4.2.1.3).
GN Name=acnA; Synonym=acn; OrderedLocusNames=BMAA1868;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bacon N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauber H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bacon N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauber H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nieman W.C., Deshazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M., Dodson R.J.,
RA Daugherty S.C., Davidsen T.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Madupu A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Durkin A.S., Nelson M.L., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR ENBL; CP000011; AAU45605.1; -; Genomic DNA.
DR TIGR; BMAA1868; -;
DR GO; GO:0003994; F:aconitase hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR00573; Aconitase C.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome; Lysase.
SQ SEQUENCE 864 AA; 93962 MW; 39963B9C243B96C2 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 406 CTNTSNP 413

RESULT 28
Q63NU1_BURPS PRELIMINARY; PRT; 864 AA.
ID Q63NU1_BURPS
AC Q63NU1;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Aconitase hydratase 1 (EC 4.2.1.3).
GN Name=acnA; Synonym=acn; OrderedLocusNames=BPS0208;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bacon N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauber H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bacon N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauber H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nieman W.C., Deshazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M., Dodson R.J.,
RA Daugherty S.C., Davidsen T.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Madupu A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Durkin A.S., Nelson M.L., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
```

```
CC
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BX571966; CAH37653.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00694; Aconitase_C; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome; Lyase.
KW SEQUENCE 864 AA; 93904 MW; 6C286C89554F6CB1 CRC64;
SQ
Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413
|||||
RESULT 29
Q8PBT6_XANCP PRELIMINARY; PRT; 864 AA.
AC Q8PBT6;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Aconitate hydratase 1.
GN Nameaena;
OS Xanthomonas campestris pv. campestris.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A.,
RA Almeida N.F., Jr., Alves L.M.C., do Amaral A.M., Bertolini M.C.,
RA Camargo L.E.A., Camarotte G., Cannavan F., Cardozo J., Chambergo F.,
RA Ciapina L.P., Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R.,
RA El-Dorry H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C.,
RA Ferro M.I.T., Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE012199; AAM40332.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
```

```
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome.
KW SEQUENCE 864 AA; 94059 MW; 8CC8BBBA4E1696BF CRC64;
SQ
Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413
|||||
RESULT 30
Q3F693_9BURK PRELIMINARY; PRT; 865 AA.
AC Q3F693;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=BambDRAFT1678;
OS Burkholderia ambifaria AMMD.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=339670;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Burkholderia ambifaria
RT AMMD."
RL Submitted (AUG-2005) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia ambifaria
RT AMMD."
RL Submitted (AUG-2005) to the EMBL/GenBank/DBSJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJU10100010; EAO44596.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
```

```
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
SQ SEQUENCE 865 AA; 94086 MW; A5EB8AD9789176A5 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 407 CTNTSNPR 414

RESULT 31
Q3RQY1_RALME PRELIMINARY; PRT; 865 AA.
AC Q3RQY1;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=RmetDRAFT_4095;
OS Ralstonia metallidurans (strain CH34).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Cupriavidus.
OX NCBI_TaxID=266264;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Ralstonia
RT metallidurans CH34.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Ralstonia metallidurans
RT CH34.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AA010300006; EAN50810.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHY.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase; _plasmid; Transmembrane.
SQ SEQUENCE 865 AA; 93731 MW; 9D59EB071A2D2004 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNPR 413

RESULT 32
Q8XTI8_RALSO PRELIMINARY; PRT; 865 AA.
AC Q8XTI8;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE PROBABLE ACONITATE HYDRATASE 1 TRANSMEMBRANE PROTEIN (EC 4.2.1.3).
GN Name=acnA2; OrderedLocusNames=RSP0120; ORFNames=RS03002;
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=GM11009;
RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AL646076; CAD17271.1; -; Genomic DNA.
DR Biocyc; RSOL305:RSP0120-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHY.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase; _plasmid; Transmembrane.
SQ SEQUENCE 865 AA; 93731 MW; 9D59EB071A2D2004 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNPR 413
```



```

RT 3.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ANA-3;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Shewanella sp. ANA-3.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALH01000012; EAP19913.1; -: Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 867 AA; 94752 MW; 66A2586E7C2C843C CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 410 CTNTSNPR 417

RESULT 35
Q3P203 9GAMM PRELIMINARY; PRT; 867 AA.
AC Q3P203;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=SDendRAFT_0483;
OS Shewanella denitrificans OS217.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318161;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella
RT denitrificans OS-217.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella denitrificans
RT OS-217.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALU01000010; EAN70542.1; -: Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
```

```

RT 3.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ANA-3;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Shewanella sp. ANA-3.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALH01000012; EAP19913.1; -: Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 867 AA; 94742 MW; 06C78A219E1371D7 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 410 CTNTSNPR 417

RESULT 35
Q3P203 9GAMM PRELIMINARY; PRT; 867 AA.
AC Q3P203;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=SDendRAFT_0483;
OS Shewanella denitrificans OS217.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318161;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella
RT denitrificans OS-217.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella denitrificans
RT OS-217.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALU01000010; EAN70542.1; -: Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
```

```
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 867 AA; 94825 MW; 9567049521DCB0A0 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. NO. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNER 413

RESULT 36
Q4J5X3 AZOVI
ID Q4J5X3 AZOVI PRELIMINARY; PRT; 867 AA.
AC Q4J5X3;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Aconitase hydratase, C-terminal; Aconitase hydratase, N-terminal.
DE ORFNames=AvindRAFT_5705;
GN Azotobacter vinelandii AVOP.
OS Azotobacter vinelandii AVOP.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=322710;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Azotobacter vinelandii
RT AVOP."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Azotobacter vinelandii
RT AVOP."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RA DOE Joint Genome Institute;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAU03000001; EAM07018.1; -; Genomic DNA.
DR GO:0016836; F:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0016829; F:lyase activity; IEA.
DR GO:0005506; F:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0005506; F:lyase activity; IEA.
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 867 AA; 94721 MW; D33BB751213AB8F4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. NO. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 413 CTNTSNER 420

RESULT 37
Q8EJW3 SHEON
ID Q8EJW3 SHEON PRELIMINARY; PRT; 867 AA.
AC Q8EJW3;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 21-FEB-2006, entry version 13.
DE Aconitase hydratase 1.0343;
GN Name=acnA; ORFNames=SO_0343;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MR-1;
RG MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Methe B.A.,
RA Clayton R.A., Meyer I., Tsapin A., Scott J., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Imbraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Neallson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AE014299; AAN53428.1; -; Genomic DNA.
DR BioCyc; SONE211586:SO0343-MONOMER; -;
DR GO:0016836; F:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0005506; F:lyase activity; IEA.
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
Complete proteome.
KW
```

```

SQ SEQUENCE 867 AA; 94596 MW; DA7DBA64A0864A1F CRC64;
Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 410 CTNTSNPR 417

RESULT 38
Q3CT85 ALTAT PRELIMINARY; PRT; 868 AA.
AC Q3CT85;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=PatIDRAFT_3346;
OS Pseudoalteromonas atlantica T6c.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudoalteromonadales; Pseudoalteromonas.
OX NCBI_TaxID=342610;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T6c;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Pseudoalteromonas
RT atlantica T6c.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T6c;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Pseudoalteromonas
RT atlantica T6c.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
-!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAKP0100011; EAO68319.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 868 AA; 95594 MW; 764CAC26C31D9B5E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413

RESULT 39
Q48017 COLP3 PRELIMINARY; PRT; 868 AA.
AC Q48017;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Putative 2-methyl-cis-aconitic acid hydratase.
GN OrderedLocusNames=CPS_2820;

```

```

OS Colwellia psychrerythraea (strain 34H / ATCC BAA-681) (Vibrio
OS psychrerythrus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Colwelliaceae; Colwellia.
OX NCBI_TaxID=167879;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043709; DOI=10.1073/pnas.0504766102;
RA Methe B.A., Nelson K.E., Deming J.W., Momen B., Melamud E., Zhang X.,
RA Moutl J., Madupu R., Nelson W.C., Dodson R.J., Brinkac L.M.,
RA Daugherty S.C., Durkin A.S., DeBoy R.T., Kolonay J.F., Sullivan S.A.,
RA Zhou L., Davidson T.M., Wu M., Huston A.L., Lewis M., Weaver B.,
RA Weidman J.F., Khouri H., Utterback T.R., Feldblyum T.V., Fraser C.M.;
RT "The psychrophilic lifestyle as revealed by the genome sequence of
RT Colwellia psychrerythraea 34H through genomic and proteomic
RT analyses.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:10913-10918 (2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP0000083; AAZ25694.1; -; Genomic DNA.
DR TIGR; CPS_2820;
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 94999 MW; FDC0F33111617B0E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 411 CTNTSNPR 418

RESULT 40
Q7NWD6 CHRVO PRELIMINARY; PRT; 868 AA.
AC Q7NWD6;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Aconitate hydratase (EC 4.2.1.3).
GN Name=acnA2; OrderedLocusNames=CV2054; ORFNames=CV_2054;
OC Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=5336;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1932124100;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA Camargo A.A., Cardoso D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chusere L.M.O.,

```

RA Creczynski-Passa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantiatti F., Farias I.P., Felipe M.S.A., Ferrari L.P., Ferro J.A.,
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA Grattapaglia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Peralha M., Pinto L.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seanez H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA Souza R.C., Steffens M.B.R., Steindl M., Teixeira S.R., Urmenyi T.,
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of Chromobacterium violaceum reveals
RT remarkable and exploitable bacterial adaptability.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AE016825; AA059726.1; -; Genomic_DNA.
DR BioCyc; CVI0243365; CV2054-MONOMER; -;
DR GO; GO:0003994; F:aconitase hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase;
SQ SEQUENCE 868 AA; 94648 MW; D0FB936453F66F95 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
| | | | |
DB 411 CTNTSNPR 418

RESULT 41
Q915E4 PSBAE
ID Q915E4 PSBAE PRELIMINARY; PRT; 868 AA.
AC Q915E4;
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2001, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Probable aconitase hydratase.
GN OrderedLocusNames=PA0794;
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 15692 / PA01.
RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H. Jr., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AE004514; AAG04183.1; -; Genomic_DNA.
DR BioCyc; PAER287; PA0794-MONOMER; -;
DR LinkHub; Q915E4; -;
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 94891 MW; 5A6E0C9D30EC328F CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
| | | | |
DB 411 CTNTSNPR 418

RESULT 42
Q9JUT05 NEIMA
ID Q9JUT05 NEIMA PRELIMINARY; PRT; 868 AA.
AC Q9JUT05;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE Aconitase hydratase (EC 4.2.1.3).
GN Name=acnA; OrderedLocusNames=NMA2052;
OS Neisseria meningitidis serogroup A.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919; DOI=10.1038/35006655;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K.L., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491";
RL Nature 404:502-506(2000).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AL162758; CAB85270.1; -; Genomic_DNA.
DR PIR; H81775; H81775.
DR BioCyc; NMN65699; NMA2052-MONOMER; -;
DR GO; GO:0003994; F:aconitase hydratase activity; IEA.

```

DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dHYP.
DR InterPro: IPR012084; AcoLysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 868 AA; 95030 MW; 0D7FE359E772F9A8 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 412 CTNTSNR 419

RESULT 43
Q9K0X3 NEIMB PRELIMINARY; PRT; 868 AA.
AC Q9K0X3;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DE Aconitate hydratase 1 (EC 4.2.1.3).
GN Name: acaA; OrderedLocNames: NM004033;
OS Neisseria meningitidis serogroup B.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MC58 / Serogroup B.
RX MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecho A., Parksey D.S., Blair E.,
RA Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Maignani V., Pizzo M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL: AF002098; AAF40871.1; -, Genomic_DNA.
DR PIR: C81200; C81200.
DR TIGR: NM00433; -.
DR BioCyc: NMEN491:NMEN491-33-MONOMER; -.
DR GO: GO:0003994; F:aconitate hydratase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dHYP.
DR InterPro: IPR012084; AcoLysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.

```

```

DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 868 AA; 95037 MW; 1179C795F538B037 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 412 CTNTSNR 419

RESULT 44
Q9KSC0 VIBCH PRELIMINARY; PRT; 868 AA.
AC Q9KSC0;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DE Aconitate hydratase 1.
GN OrderedLocNames=VCL338;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL: AF004213; AAF94496.1; -, Genomic_DNA.
DR PIR: B82213; B82213.
DR TIGR: VCI338; -.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dHYP.
DR InterPro: IPR012084; AcoLysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 95073 MW; E8837B65E9FDED CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 415 CTNTSNR 422

```

```

RESULT 45
Q31H91_PSEHT PRELIMINARY; PRT; 869 AA.
AC Q31H91;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Aconitate hydratase 1 (EC 4.2.1.3).
GN Name=acnA; OrderedLocNames=PSHA1773; ORFNames=PSHA1773;
OS Pseudocalteromonas haloplanktis (strain TAC 125).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudoalteromonadaceae; Pseudoalteromonas.
OX NCBI_TaxID=326442;
RN [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16169927; DOI=10.1101/gr.4126905;
RA Medigue C., Krin E., Pascal G., Barbe V., Bernsel A., Bertin P.N.,
RA Cheung F., Cruveiller S., D'Amico S., Duilio A., Fang G., Feller G.,
RA Ho C., Mangerot S., Marino G., Nilsson J., Parrilli E., Rocha E.P.C.,
RA Rouy Z., Sekowska A., Tutino M.L., Vallenet D., von Heijne G.,
RA Danchin A.;
RT "Coping with cold: the genome of the versatile marine Antarctic
RT bacterium Pseudocalteromonas haloplanktis TAC125.";
RL Genome Res. 15:1325-1335(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: CR954246; CAI86845.1; -: Genomic DNA.
DR GO: GO:0003994; F:aconitate hydratase activity; IEA.
DR GO: GO:0005066; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 869 AA; 95141 MW; EB6A02459E658899 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 410 CTNTSNPR 417

RESULT 46
Q937N8_RALEU PRELIMINARY; PRT; 869 AA.
AC Q937N8;
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Probable methyl-cis-aconitic acid hydratase.
GN Name=acnM;
OS Ralstonia eutropha (Alcaligenes eutrophus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Cupriavidus.
OX NCBI_TaxID=106590;
RN [1]
NUCLEOTIDE SEQUENCE.
RC STRAIN=HF39;

```

```

RX MEDLINE=21387530; PubMed=11495997;
RA Bramer C.O., Steinbuechel A.;
RT "The methylicitric acid pathway in Ralstonia eutropha: new genes
RT identified involved in propionate metabolism.";
RL Microbiology 147:2203-2214(2001).
RN [2]
NUCLEOTIDE SEQUENCE.
RC STRAIN=HF39;
RA Bramer C.O., Steinbuechel A.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AF325554; AAL03990.1; -: Genomic DNA.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 869 AA; 94726 MW; DA11CA78DD0C9710 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 411 CTNTSNPR 418

RESULT 47
Q2KW75_BORAV PRELIMINARY; PRT; 869 AA.
AC Q2KW75;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Aconitate hydratase 2 (EC 4.2.1.3).
GN Name=acnA4; ORFNames=BAV2732;
OS Bordetella avium 197N.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Alcaligenaceae; Bordetella.
RN [1]_TaxID=360910;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=197N;
RA Sebaihia M.;
RT "The genome sequence of the poultry pathogen Bordetella avium, and
RT genomic comparisons with related species infecting mammals.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AM167904; CAJ50343.1; -: Genomic DNA.
DR Lyase.
SQ SEQUENCE 869 AA; 95233 MW; 0BC85A015BE50385 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8

```



```

CC      preliminary data.
CC      -----
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      -----
DR      EMBL; AALB01000004; EAO95942.1; -; Genomic DNA.
SQ      SEQUENCE 871 AA; 95191 MW; 470199BA8368C1F3 CRC64;

Query Match      70.5%; Score 43; DB 2; Length 871;
Best Local Similarity 87.5%; Pred. NO. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTNTDNPR 8
      |||||
Db      414 CTNTSNPR 421

Search completed: June 5, 2006, 12:54:32
Job time : 121.671 secs

```


GenCore version 5.1.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 104.268 Seconds
(without alignments)
2293.354 Million cell updates/sec

Title: US-10-645-659A-5
Perfect score: 2728
Sequence: 1 MEVLILLVLLAVPPRRTRAE.....LPAFSYGFYVRNKAIAICI 523

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*
9: Geneseq2005s:*
10: Geneseq2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2728	100.0	523	5 ABB07814	Chicken h
2	2728	100.0	523	7 ABW02017	Abw02017 Chicken h
3	2728	100.0	523	8 ADR88211	Adr88211 Chicken h
4	2728	100.0	523	8 ADT78178	Adt78178 Chicken h
5	2728	100.0	523	9 ADY27037	Ady27037 Chicken h
6	2728	100.0	523	9 AEA42427	Aea42427 Chicken h
7	1717	62.9	527	5 ABB07815	Abb07815 Chicken s
8	1717	62.9	527	7 ABW02018	Abw02018 Chimeric
9	1711	62.7	527	8 ADO63825	Ado63825 Chimeric
10	1711	62.7	527	8 ADO63826	Ado63826 Chimeric
11	1705	62.5	527	8 ADO63827	Ado63827 Chimeric
12	1653.5	60.6	543	8 ADO63832	Ado63832 Human hep
13	1650	60.5	545	6 ADP56822	Adp56822 Human hep
14	1650	60.5	545	7 ADE16012	Adel6012 G-coupled
15	1650	60.5	545	8 ADL93951	Adl93951 Human G-c
16	1648.5	60.4	543	2 AAY17082	Aay17082 Human hep
17	1648.5	60.4	543	4 AAB86206	Aab86206 Human hep
18	1648.5	60.4	543	7 ADD18950	Add18950 Human dis
19	1648.5	60.4	543	8 ADK52086	Adk52086 Human ato
20	1648.5	60.4	543	8 ADM48759	Adm48759 Human hpa
21	1648.5	60.4	543	8 ADM05074	Adm05074 Antipori
22	1648.5	60.4	543	8 ADN04902	Adn04902 Antipori
23	1648.5	60.4	543	8 ADQ80372	Adq80372 Heparanas

24	1648.5	60.4	543	8 ADR88210	Adr88210 Human pre
25	1648.5	60.4	543	8 ADP25079	Adp25079 PRO polyep
26	1648.5	60.4	543	8 ADT78177	Adt78177 Human hep
27	1648.5	60.4	543	9 ADY27036	Ady27036 Human hep
28	1648.5	60.4	543	9 AEA42426	Aea42426 Human hep
29	1648.5	60.4	588	2 AAY30124	Aay30124 A human h
30	1645.5	60.3	543	2 AAY02345	Aay02345 A human h
31	1645.5	60.3	543	3 AAY57590	Aay57590 Human hep
32	1645.5	60.3	543	3 AAB08849	Aab08849 Amino aci
33	1645.5	60.3	543	3 AAY52990	Aay52990 Human hep
34	1645.5	60.3	543	4 AAY97635	Aay97635 Human hep
35	1645.5	60.3	543	5 ABB07813	Abb07813 Human hep
36	1645.5	60.3	543	7 ADG88800	Adg88800 Human hpa
37	1645.5	60.3	543	8 ADL16379	Adl16379 Human hep
38	1645.5	60.3	543	8 ADM48716	Adm48716 Human hpa
39	1645.5	60.3	543	9 AEA42466	Aea42466 Human hep
40	1645.5	60.3	543	10 AEE96848	Aee96848 Human hep
41	1645.5	60.3	592	2 AAY02346	Aay02346 A human h
42	1645.5	60.3	592	3 AAB08850	Aab08850 Amino aci
43	1645.5	60.3	592	7 ADG88804	Adg88804 Human SK-
44	1645.5	60.3	592	8 ADL16383	Adl16383 Human hep
45	1645.5	60.3	592	8 ADM48720	Adm48720 Human SK-

ALIGNMENTS

RESULT 1
ABB07814
ID ABB07814 standard; protein; 523 AA.
XX
AC ABB07814;
XX
DT 03-JUL-2002 (first entry)
XX
DE Chicken heparanase sequence.
XX
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; chicken.
XX
OS Gallus gallus.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "putative signal peptide"
FT /note= "mature protein"
XX
FN US2002034810-A1.
XX
PD 21-MAR-2002.
XX
PF 16-AUG-2001; 2001US-00930218.
XX
PR 20-SEP-2000; 2000US-00666390.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX
DR WPI; 2002-338926/37.
XX
DR N-PSDB; ABL40748.
XX
PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
PT to treat various heparin-related disorders and the signal peptide is
PT useful in production of membrane-targeted or secreted recombinant
PT proteins.
PS Claim 19; Fig 1b; 39pp; English.
XX
CC The invention relates to an isolated avian and reptile nucleic acid,
CC encoding a polypeptide with heparanase catalytic activity. The signal
CC peptide of the nucleic acid can be used to express membrane-associated or

Db 361 MRQVSFGAGSYHLVDAGFKPLPDYWLSLYKRLVGTTRVLAQSVQADARRPRVYLHCTNP 420
 QY 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYLLPHGKDSLREVLNGRLL 480
 Db 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYLLPHGKDSLREVLNGRLL 480
 QY 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFVIRNAKAIACI 523
 Db 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFVIRNAKAIACI 523

RESULT 3

ADR88211

ID ADR88211 standard; protein; 523 AA.

AC ADR88211;

XX ADR88211;

DT 18-NOV-2004 (first entry)

XX Chicken heparanase.

XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;

KW autoimmune disorder; cancer; angiogenesis; metastatic disease;

KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;

KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;

KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;

KW chicken; heparanase; enzyme.

XX Gallus gallus.

OS

XX

FH

FT Peptide

FT 1: 19

FT /label= Signal_peptide

FT Protein

FT 20: 523

FT /label= Mature_heparanase

XX US2004170631-A1.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC of a patient, the tissue expressing heparanase. The method comprises
 CC providing a complex of a drug directly or indirectly linked to an anti-
 CC heparanase antibody, and administering the complex to the patient. In the
 CC targeted drug delivery, the antibody comprises an epitope of a heparanase
 CC capable of specifically binding to at least one epitope of a heparanase
 CC protein. The composition and methods of the invention are useful for
 CC diagnosing, preventing or treating conditions associated with heparanase
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
 CC proliferation, invasion of circulating tumour cells and metastatic
 CC disease), for purifying heparanase, or for developing drugs for those
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
 CC multiple myeloma. The solid cancer is selected from tumours in lip and
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
 CC sarcoma. The present sequence is chicken heparanase.

XX Sequence 523 AA;

Query Match 100.0%; Score 2728; DB 8; Length 523;
 Best Local Similarity 100.0%; Pred. No. 1.5e-267;
 Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLVLLLVLLAVPPRTAEQLGLRBPFGAVSPAFSLTLTDASLARDPRFVALLRHPKL 60
 Db 1 MLVLLLVLLAVPPRTAEQLGLRBPFGAVSPAFSLTLTDASLARDPRFVALLRHPKL 60
 QY 61 HTLASGLSPGFLREGGTTSTDELIENPNKDSWEEKVLSFQAKDYCEAWPSFVVPKLLL 120
 Db 61 HTLASGLSPGFLREGGTTSTDELIENPNKDSWEEKVLSFQAKDYCEAWPSFVVPKLLL 120
 QY 121 TQWPLQEKLLLAHSHWKKHKNTTITRSTLDTLHTFASSSGPRLVFGLNALLRRAGLQWDS 180
 Db 121 TQWPLQEKLLLAHSHWKKHKNTTITRSTLDTLHTFASSSGPRLVFGLNALLRRAGLQWDS 180
 QY 181 SNAKQLLYCAQRSYNI SWELGNEPNSFRKKSIGCIDGFLGRDFVHLRQLLSOHPLYRH 240
 Db 181 SNAKQLLYCAQRSYNI SWELGNEPNSFRKKSIGCIDGFLGRDFVHLRQLLSOHPLYRH 240
 QY 241 AELYGLDVGGPRKHTQHLRSFMKSGGKADSVTWHYHYNGRSATREDFISPEVLDSFA 300
 Db 241 AELYGLDVGGPRKHTQHLRSFMKSGGKADSVTWHYHYNGRSATREDFISPEVLDSFA 300
 QY 301 TAIHDLVIGIVEATVPKKVWLGTSAYGGAPOLSNITYVAGFMWLDKGLAARIGIDV 360
 Db 301 TAIHDLVIGIVEATVPKKVWLGTSAYGGAPOLSNITYVAGFMWLDKGLAARIGIDV 360
 QY 361 MRQVSFGAGSYHLVDAGFKPLPDYWLSLYKRLVGTTRVLAQSVQADARRPRVYLHCTNP 420
 Db 361 MRQVSFGAGSYHLVDAGFKPLPDYWLSLYKRLVGTTRVLAQSVQADARRPRVYLHCTNP 420
 QY 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYLLPHGKDSLREVLNGRLL 480
 Db 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYLLPHGKDSLREVLNGRLL 480
 QY 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFVIRNAKAIACI 523
 Db 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFVIRNAKAIACI 523

Targeted drug delivery to a heparanase-expressing tissue of a patient,
 useful for treating heparanase-associated conditions such as inflammation
 or cancer, comprises administering a drug and an anti-heparanase antibody
 complex.

Claim 2; SEQ ID NO 5; 58pp; English.

The invention relates to a method of targeted drug delivery to a tissue

RESULT 4

ADT78178
 ID ADT78178 standard; protein; 523 AA.
 AC
 AC ADT78178;
 DT 13-JAN-2005 (first entry)
 DE
 DE Chicken heparanase protein.
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnary; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; chicken.
 XX
 OS Gallus gallus.
 XX
 FH Key Location/Qualifiers
 FH Binding-site 136..141
 FT /note= "Putative heparin binding site"
 FT Binding-site 251..257
 FT /note= "Putative heparin binding site"
 FT Binding-site 406..413
 FT /note= "Putative heparin binding site"
 FT
 FT US2004213789-A1.
 PN
 XX
 PD 28-OCT-2004.
 XX
 XX 22-AUG-2003; 2003US-00645659.
 PF
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX
 XX (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (GELD/) GELDER J M V.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX
 PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;
 XX
 DR WPI; 2004-774790/76.

XX
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX
 PS Claim 5; SEQ ID NO 5; 68pp; English.

XX
 CC The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related

CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents chicken heparanase.

XX
 SQ Sequence 523 AA;

Query Match 100.0%; Score 2728; DB 8; Length 523;

Best Local Similarity 100.0%; Pred. No. 1.5e-267; Indels 0; Gaps 0;

Matches 523; Conservative 0; Mismatches 0;

QY	1	MLVLLLVLLVAVPPRTAEQLGLREPIGAVSPAFSLTLDSIARDPRFVALLRHPKL	60
DB	1	MLVLLLVLLVAVPPRTAEQLGLREPIGAVSPAFSLTLDSIARDPRFVALLRHPKL	60
QY	61	HTLASGLSPGLRPGTSTDFLFPNPKDSTWEEKVLSFOAKDVCCEWPSFAVVPKLLL	120
DB	61	HTLASGLSPGLRPGTSTDFLFPNPKDSTWEEKVLSFOAKDVCCEWPSFAVVPKLLL	120
QY	121	TQWPLQEKLLLAHESWKKHKNNTITRSTLDLHTFASSSGFRLVFGNALRRAGLQWDS	180
DB	121	TQWPLQEKLLLAHESWKKHKNNTITRSTLDLHTFASSSGFRLVFGNALRRAGLQWDS	180
QY	181	SNAKQLLGYCAQRSYNIISWELGNEPNSFRKKSIGCIDGFGQGRDFVHLRQLLSOHPLYRH	240
DB	181	SNAKQLLGYCAQRSYNIISWELGNEPNSFRKKSIGCIDGFGQGRDFVHLRQLLSOHPLYRH	240
QY	241	AELYGLDVGQPRKHTQHLRSFMKSGGKAIDSVTWHHYNGRSATREDFLSPEVLSFA	300
DB	241	AELYGLDVGQPRKHTQHLRSFMKSGGKAIDSVTWHHYNGRSATREDFLSPEVLSFA	300
QY	301	TAIHDVLGIVEATVPGKKVWLGTSAYGGAPQLSNTYVAGFMWLDKLGIAARRGIDV	360
DB	301	TAIHDVLGIVEATVPGKKVWLGTSAYGGAPQLSNTYVAGFMWLDKLGIAARRGIDV	360
QY	361	MRQVSFGAGSYHLVDAGFKPLPDYWLSSLKYRLVGTQVLOASVEQADARRPRVYLHCTNP	420
DB	361	MRQVSFGAGSYHLVDAGFKPLPDYWLSSLKYRLVGTQVLOASVEQADARRPRVYLHCTNP	420
QY	421	RHPKYREGDVTLPALNLSNVTQSLQLPKQLWSKSDVQYLLPHGKDSILSREVQNGRL	480
DB	421	RHPKYREGDVTLPALNLSNVTQSLQLPKQLWSKSDVQYLLPHGKDSILSREVQNGRL	480
QY	481	QMVDDETLPALHEMALAPGSTLGLPAFSYGFYVIRNAKATACI	523
DB	481	QMVDDETLPALHEMALAPGSTLGLPAFSYGFYVIRNAKATACI	523

RESULT 5

ADY27037
 ID ADY27037 standard; protein; 523 AA.

XX
 AC ADY27037;

DT 05-MAY-2005 (first entry)

DE Chicken heparanase protein.

KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.

XX
 OS Gallus gallus.

XX WQ2005016227-A2.
XX
XX
XX 24-FEB-2005.
XX
XX 12-AUG-2004; 2004WO-IL000744.
XX
XX 14-AUG-2003; 2003US-0494800P.
XX 13-JAN-2004; 2004US-0535492P.
XX
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX
XX Van-Gelder JM, Miron D;
XX
XX WPI; 2005-182203/19.
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
XX diseases (e.g. cancer, inflammation, cardiovascular diseases,
XX neurological diseases or viral diseases) comprises modulating heparanase
XX activation.
XX
XX Disclosure; SEQ ID NO 9; 211pp; English.
XX
XX The invention relates to a method of regulating heparanase activity in a
XX tissue or regulating a biological process depending at least in part on
XX heparanase activity comprising modulating heparanase activation. The
XX invention also relates to methods of treating a heparanase- or heparin
XX binding protein-associated disease or disorder in a subject, a
XX pharmaceutical composition for use in the treatment of a heparanase-
XX associated disease or disorder comprising a therapeutic amount of an
XX agent capable of modulating heparanase activation and a pharmaceutical
XX carrier or diluent, a method of identifying a protease activator of
XX heparanase, a protease substrate mimetic comprising a peptide
XX representing a subset or all substrate residues or cleavage sites of
XX human heparanase or an equivalent non-human heparanase, a method of
XX producing active heparanase and a method of modulating an adhesion
XX activity of heparanase. The composition and methods are useful for
XX modulating heparanase activation and for treating heparanase-associated
XX diseases or disorders such as cancer, inflammation, cardiovascular
XX diseases, neurological diseases or viral infections. This sequence
XX represents a chicken heparanase protein used in the scope of the
XX invention.
XX
XX Sequence 523 AA;

Query Match 100.0%; Score 2728; DB 9; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.5e-267;
Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLVLLLVLLVAVPPRTAEIQLGLRPIGAVSPAFSLTLDSARDPRFVALLRHPKL 60
DB 1 MLVLLLVLLVAVPPRTAEIQLGLRPIGAVSPAFSLTLDSARDPRFVALLRHPKL 60
QY 61 HTLASGLSPGLRPGGTSTDFLIPNPKDSTWEEKVSEFOAKDVCWAPSFVAVPKLL 120
DB 61 HTLASGLSPGLRPGGTSTDFLIPNPKDSTWEEKVSEFOAKDVCWAPSFVAVPKLL 120
QY 121 TQVLOEKLLAEHSWKHKNTTITRSTDLTHTFASSSGRLVFGNLNLRAGLWDS 180
DB 121 TQVLOEKLLAEHSWKHKNTTITRSTDLTHTFASSSGRLVFGNLNLRAGLWDS 180
QY 181 SNAKQLLGCAQRSYNTSWELGNEFPNFRKKSIGCIDGFQGRDFVHLRQLLSQHPLYRH 240
DB 181 SNAKQLLGCAQRSYNTSWELGNEFPNFRKKSIGCIDGFQGRDFVHLRQLLSQHPLYRH 240
QY 241 AELYGLDVGPQRKHTQHLRSFMKSGKAIDSVTWHYVNGRSATREDFLSPEVLSFA 300
DB 241 AELYGLDVGPQRKHTQHLRSFMKSGKAIDSVTWHYVNGRSATREDFLSPEVLSFA 300
QY 301 TAIHDLVGIVEATVPGKKVWLGTSYAGGAPQLSNTYVAGFWMLDKLGLAARGIDV 360
DB 301 TAIHDLVGIVEATVPGKKVWLGTSYAGGAPQLSNTYVAGFWMLDKLGLAARGIDV 360

QY 361 MRQVSFGAGSYHLVDAGFKPLPDYWLSTLLYKRLVGRVLOASVEQADARRPRVYLHCTNP 420
DB 361 MRQVSFGAGSYHLVDAGFKPLPDYWLSTLLYKRLVGRVLOASVEQADARRPRVYLHCTNP 420
QY 421 RHPKYREGDVTFLFALNLSNVTQSLQPKQLMSKSDVQYLLPHGKDSILSREVQLNGRL 480
DB 421 RHPKYREGDVTFLFALNLSNVTQSLQPKQLMSKSDVQYLLPHGKDSILSREVQLNGRL 480
QY 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGFGVYIRNAKAIACI 523
DB 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGFGVYIRNAKAIACI 523
RESULT 6
AEA42427
ID AEA42427 standard; protein; 523 AA.
XX
XX AEA42427;
XX
XX 28-JUL-2005 (first entry)
XX
XX Chicken heparanase epitope peptide SEQ ID NO:5.
XX
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
XX antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
XX inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
XX angiogenesis disorder; cancer; tumor; metastasis.
XX
XX Gallus gallus.
XX
XX AU2004201462-A1.
XX
XX 06-MAY-2004.
XX
XX 08-APR-2004; 2004AU-00201462.
XX
XX 08-APR-2004; 2004AU-00201462.
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
XX Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
XX Feinstein E;
XX
XX WPI; 2005-173343/19.
XX
XX Novel isolated antibody capable of specifically binding to epitope of
XX heparanase protein, useful for preventing and treating heparanase-related
XX disorder such as inflammatory disorder, scars, autoimmune conditions or
XX angiogenesis.
XX
XX Claim 2; SEQ ID NO 5; 260pp; English.
XX
XX The invention relates to an isolated antibody or its portion (I) capable
XX of specifically binding to an epitope of a heparanase protein. Also
XX described: (1) a cell line (II) for producing a monoclonal antibody or
XX its portion, comprising a cell line for producing (I); (2) a
XX pharmaceutical composition comprising (I) and a carrier; and (3) an
XX affinity medium (III) for binding human heparanase polypeptides,
XX comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
XX useful for treating a subject suffering from a pathological condition,
XX which involves administering (I) to the subject. (I) is useful for
XX preventing and treating heparanase-related disorder or condition chosen
XX from inflammatory disorder, wound, scar, vasculopathy, autoimmune
XX condition, angiogenesis, cell proliferation, cancerous condition, tumor
XX cell proliferation, invasion of circulating tumor cells and metastatic
XX disease. (I) is useful for detecting the presence of heparanase
XX polypeptide in a sample. (I) is useful for detecting heparanase-related
XX disease or condition in a subject such as vertebrate, preferably mammal
XX e.g., human. The heparanase-related disorder or condition further
XX includes renal disease or disorder chosen from diabetic nephropathy,
XX glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome

CC and renal cell carcinoma. The present sequence represents chicken
CC heparanase, which is used in the exemplification of the present
CC invention.

XX Sequence 523 AA;

Query Match 100.0%; Score 2728; DB 9; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.5e-267;
Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLVLLLVLLVAVPPRRTAELQGLRPIGAVSPAFSLTLTDLASLARDPRFVALLRHPKL 60
DB 1 MLVLLLVLLVAVPPRRTAELQGLRPIGAVSPAFSLTLTDLASLARDPRFVALLRHPKL 60
QY 61 HTLASGLSPGFLRFGGTSTDFLIPNPKDSTWEEKVSEFQAKDVCCEAMPSPFAVVPKLL 120
DB 61 HTLASGLSPGFLRFGGTSTDFLIPNPKDSTWEEKVSEFQAKDVCCEAMPSPFAVVPKLL 120
QY 121 TQWPLEKLLLAHSHWKKHKNNTITRSTLDILHTFASSSGFRLVFGNLALLRAGLOWDS 180
DB 121 TQWPLEKLLLAHSHWKKHKNNTITRSTLDILHTFASSSGFRLVFGNLALLRAGLOWDS 180
QY 181 SNAKQLLGYCAORSYINISWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLROLLSOHPLYRH 240
DB 181 SNAKQLLGYCAORSYINISWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLROLLSOHPLYRH 240
QY 241 AELYGLDVGQPRKHTQHLLRSFMKSGKAIDSVTWHYVNGRSATREDFLSPEVLDSPA 300
DB 241 AELYGLDVGQPRKHTQHLLRSFMKSGKAIDSVTWHYVNGRSATREDFLSPEVLDSPA 300
QY 301 TAIHDVLGIVEATVPGKKVWLGTGSAYGGAQPOLSNYYVAGFMWLDKGLAARRGIDVV 360
DB 301 TAIHDVLGIVEATVPGKKVWLGTGSAYGGAQPOLSNYYVAGFMWLDKGLAARRGIDVV 360
QY 361 MRQVSFGAGSYHLVDAGKPLPDYWLSSLVKRLVGTGRVLOASVEQADARRPRVYLHCTNP 420
DB 361 MRQVSFGAGSYHLVDAGKPLPDYWLSSLVKRLVGTGRVLOASVEQADARRPRVYLHCTNP 420
QY 421 RHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVQYLLPHGKDSILSREVLNGRL 480
DB 421 RHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVQYLLPHGKDSILSREVLNGRL 480
QY 481 QMWDDTLPALHEMALAPGSTLGLPAFSYGFYVIRNAKAIACI 523
DB 481 QMWDDTLPALHEMALAPGSTLGLPAFSYGFYVIRNAKAIACI 523

RESULT 7
ABB07815
ID ABB07815 standard; protein; 527 AA.

XX ABB07815;

DT 03-JUL-2002 (first entry)

XX Chicken signal peptide/human heparanase chimeric protein sequence.

DE Heparanase: catalytic; cytoaratic; antiviral; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX Synthetic.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note= "chicken heparanase signal peptide"

FT Protein /note= "human heparanase mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX PF 16-AUG-2001; 2001US-00930218.
XX PR 20-SEP-2000; 2000US-00666390.
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX DR WPI; 2002-338926/37.
XX DR N-PSDB; ABL40753.
XX PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful
XX PT to treat various heparin-related disorders and the signal peptide is
XX PT useful in production of membrane-targeted or secreted recombinant
XX PT proteins.
XX PS Disclosure; Page 26-28; 39pp; English.
XX CC The invention relates to an isolated avian and reptile nucleic acid,
XX CC encoding a polypeptide with heparanase catalytic activity. The signal
XX CC peptide of the nucleic acid can be used to express membrane-associated or
XX CC secreted proteins in heterologous expression systems. The encoded
XX CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
XX CC invasion, and to intervene with pathologies associated with impaired
XX CC heparin-binding growth factors, cellular responses to heparin-binding
XX CC growth factors and cytokines, cell interaction with plasma lipoproteins,
XX CC cellular susceptibility to viral, protozoa and bacterial infections or
XX CC disintegration of neurodegenerative plaques. The present sequence
XX CC represents a chicken signal peptide/human heparanase chimeric protein
XX CC sequence
XX
SQ Sequence 527 AA;

Query Match 62.9%; Score 1717; DB 5; Length 527;
Best Local Similarity 63.1%; Pred. No. 9.1e-165;
Matches 333; Conservative 83; Mismatches 106; Indels 6; Gaps 3;

QY 1 MLVLLLVLLVAVPPRRTA---ELQLGREPIGAVSPAFSLTLTDLASLARDPRFVALLR 56
DB 1 MLVLLLVLLVAVPPRRTAQDVVDLDFTEQPLHLVSPSFLSVTDANLATDPRFLILG 60
QY 57 HPKLHTLASGLSPGFLRFGGTSTDFLIPNPKDSTWEEKVSEFQA-KDVCCEAMPSPFAV 115
DB 61 SPKLRTLARGLSPAYLRFGGTKTDFLIFDPKKESTFEERSYQVQVNDICKYGSIPDV 120
QY 116 PKLLLTQWPLEKLLLAHSHWKKHKNNTITRSTLDILHTFASSSGFRLVFGNLALLRAG 175
DB 121 EEKLRUEWPYQEQLLLRHYYOKKPKNSTYGRSSVDVILTYFANCSGLDLIFGNLALLRAD 180
QY 176 LOWDSSNAKOLLGYCAORSYINISWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLROLLSOH 235
DB 181 LOWDSSNAKOLLGYCAORSYINISWELGNEPNSFRKKSIGICIDGFGOLGRDFVHLROLLSOH 239
QY 236 PLYRHAELYGLDVQPRKHTQHLLRSFMKSGKAIDSVTWHYVNGRSATREDFLSPEV 295
DB 240 STEFNAKLYGPDVGQPRKRTAKMLKSLFKAGGEVIDSVTHHHYVNGRSATREDFLNPDV 299
QY 296 LDSFATAIHVDVLGIVEATVPGKKVWLGTGSAYGGAQPOLSNYYVAGFMWLDKGLAARR 355
DB 300 LDIFISVQVQFQVESTRFPGKKVWLGTGSAYGGAQPOLSNYYVAGFMWLDKGLAARR 359
QY 356 GIDVWRQVSFGAGSYHLVDAGKPLPDYWLSSLVKRLVGTGRVLOASVEQADARRPRVYL 415
DB 360 GIEVWRQVSFGAGSYHLVDAGKPLPDYWLSSLVKRLVGTGRVLOASVEQADARRPRVYL 419
QY 416 HCTNPRHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVQYLLPHGKDSILSREVL 475
DB 420 HCTNTPRYKEGDLTLAHLNHNVTKYLRPLYPFNSKQVNDICKYGSIPDV 479
QY 476 NGRLLQWDDTLPALHEMALAPGSTLGLPAFSYGFYVIRNAKAIACI 523
DB 480 NGLTLKMVDDQTLPLPMEKPLRPGSSLGLPAFSYGFYVIRNAKAIACI 527

RESULT 8
 ABW02018
 ID ABW02018 standard; protein; 527 AA.
 AC ABW02018;
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Chimeric human-chicken heparanase protein.
 XX
 KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
 KW immunosuppressive; vulnerability; bactericide; anti-angiogenic; virucide;
 KW anticlerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.
 XX
 OS Chimeric - Gallus gallus.
 OS Chimeric - Homo sapiens.
 XX
 PN US2003180788-A1.
 XX
 XX 25-SEP-2003.
 XX
 XX 08-MAY-2003; 2003US-00431438.
 XX
 XX 20-SEP-2000; 2000US-00666390.
 PR 16-AUG-2001; 2001US-00930218.
 XX
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zoharia E;
 PI
 DR WPI; 2003-843931/78.
 DR N-PSDB; AAD63532.
 XX
 XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.
 PT
 XX Example; Page 26-28; Opp; English.
 XX
 CC The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
 CC some bacterial and protozoa infections. Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamins. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein
 XX
 SQ Sequence 527 AA;
 Query Match 62.9%; Score 1717; DB 7; Length 527;
 Best Local Similarity 63.1%; Pred. No. 9.1e-165;
 Matches 333; Conservative 83; Mismatches 106; Indels 6; Gaps 3;
 QY 1 MLVLLLVLLVLPVPRRTA---ELQLGLREPIGAVSPAFSLTLDASLARDPRFVALLR 56
 DB 1 MLVLLLVLLVLPVPRRTAQDVLDLFTQEPHLVSPFLSVITDANLATDPRFLILG 60
 QY 57 HPKHLTLASGLSPGFLRFGTSTDFLIFNPKNOSTWBEKVLSEFQA-KDVCEAWPSFAVV 115
 DB 61 SPKRLTLARGLSPAYLRFGGTKTDFLIFDPKKESTFEERSYMQSQVNQDICKYGSIPDPV 120

QY 116 PKLLLTQWPLQEKLLLAHSHWKKHNTTITBSTLDILHTFASSSGFRLVFGNLALLRRAG 175
 DB 121 EEKLRLEMPYQEQQLLREHYOKFKPNSTYSRSSVDVLYTFANCSDLIFGLNALLRAD 180
 QY 176 LQWDSSNAKQLLYGCAORSYNISWELGNEPNSFRKKGICIDGFGQLGRDFVHLRQLLSQH 235
 DB 181 LQWNSSNAQLLDYCSSKGYNISWELGNEPNSFLKADIFINGSQLGEDFIQLHKLL-RK 239
 QY 236 PLYRHABLYGLDVQOPRKHTOHLRLSRPMKSGGKAIDSVTWHYYVNGRSATREDFLSPEV 295
 DB 240 STFKNAKLYGPDVQPPRRKAKMLKFLKAGGEVIDSVTWHYYLNGRTATREDFLNPV 299
 QY 296 LDSFATAIHDLVGLVEATVPGKVMWLGTSAYGGAPQLSNYYVAGFMWLDKGLAARR 355
 DB 300 LDIFISSVQKVFQVVESTRPGKVMWLGTSAYGGAPQLSDTFAAGFMWLDKGLUSARM 359
 QY 356 GIDVVMQVSEFAGSYHLVDAGFKPLPDYWLISLYKRLVGRVLOASVEQADARRPRVYL 415
 DB 360 GIEVVMQVSEFAGSYHLVDENFDPLPDYWLISLLFKLVGTVKLVMSVQSGSKRKLRLVYL 419
 QY 416 HCTNPRHPKYREGDVTFLALNLSNVTOSLQLPKQLMWSKSDVQYLLPHGKDSILSREVQL 475
 DB 420 HCTNTDNPRYKEGDLTLVAINLHNVTKYLRPLYPSPNKQVDKYLLRPLGPHGLLSKVQL 479
 QY 476 NGRLLQWVDDDELTPALHEMALAPGSTLGLPAFSYGFYVIRNAKAIACI 523
 DB 480 NGLTLKWVDDQTLPLPMEKPLRPGSSGLGLPAFSYFFVIRNAKVAACI 527
 RESULT 9
 ADO63825
 ID ADO63825 standard; protein; 527 AA.
 XX
 AC ADO63825;
 XX
 DT 26-AUG-2004 (first entry)
 XX
 DE Chimeric heparanase mutant E225A, SEQ ID:10.
 XX
 KW Human; chicken; heparanase; heparanase-derived protein;
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
 KW adhesion disorder; IAD; leukocyte adhesion deficiency;
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
 KW vulnerability; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Gallus gallus.
 OS Synthetic.
 OS Chimeric.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..18 /note= "Chicken heparanase signal peptide"
 FT Region 19..527 /note= "Corresponds to residues 35-543 of human
 FT heparanase mutant E225A (SEQ ID NO:7)"
 FT Misc-difference 209 /note= "Ala replaces wild-type Glu (active site proton
 FT donor). Corresponds to residue 225 of human heparanase
 FT mutant E225A (SEQ ID NO:7)"
 FT Active-site 327 /note= "Active site nucleophile"
 FT
 XX WO2004048558-A2.
 XX
 XX 10-JUN-2004.
 XX
 XX 24-NOV-2003; 2003WO-IL000989.
 XX
 XX 24-NOV-2002; 2002IL-00153059.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX DR WPI; 2004-450373/42.
XX DR N-PSDB; ADO63819.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX Claim 10; SEQ ID NO 10; 128pp; English.
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a chimeric
CC protein comprising the signal peptide of chicken heparanase and residues
CC 35-543 of the human heparanase mutant E225A.
XX Sequence 527 AA;
SQ

Query Match 62.7%; Score 1711; DB 8; Length 527;
Best Local Similarity 62.9%; Pred. No. 3.7e-164;
Matches 332; Conservative 83; Mismatches 107; Indels 6; Gaps 3;

QY 1 MLVLLLLVLLAVPPRTA----ELQLGREPTGAVSPAPLSLTLDASLARDPRFVALLR 56
DB 1 MLVLLLLVLLAVPPRTAQDVLDLFFQEPHLVSPFLSVTIDANLATDPRFLILG 60
QY 57 HPKLLTASGLSPGFLRFGTSTDFLIFPNKDSWTWEEKVLSBQA-KVCEAWPSPFAV 115
DB 61 SPKRLTLARGLSPAYLRFQGTGKDFLIFDPKSTFEERSYQSNQDICKYGIIPDV 120
QY 116 PKLLLTQWLPLOEKLLIAHSWKHKHTTITRSLDILHFPASSGFLVFGIALLRRAG 175
DB 121 EEKRLLEWYQEQLLREHYQFKFNSTYSRSSVDVLYTFPANGSGLDLIFGLNALLRTAD 180
QY 176 LOWDSSNAKOLLGYCAORSYNIWELGNPNPSPKSGICIGDFOLGRDFVHLRQLLSQH 235
DB 181 LOWNSSNAQLLDYCSKGNYSWELGNPNPSPKSGICIGDFOLGRDFVHLRQLLSQH 239
QY 236 PLYRHAELYGLDVGPQRKHTQHLLRFSKSGKAIDSVTWHYVNGRSATREDFLSPVEY 295
DB 240 STFKNAKLGPDVGQPRKTAHKLKSLKAGGEVIDSVTHYVNGRTATREDFLNPDV 299
QY 296 LDSFATAIHDLVIGVATVPGKVKWLGTCGSYGGGAPQLSNTYVAGFMWLDKGLAARR 355
DB 300 LDIFISSQVKVQFVSTPRGKVKWLGTCGSYGGGAPQLSNTYVAGFMWLDKGLSARM 359
QY 356 GIDVVMRQVSFGAGSVHLVDAGPKPLPDYWLSSLKYLKLVGTRVLOASVEQADARRPRVYL 415

DB 360 GIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTGKVLMAVQSGKRRKLRVYL 419
QY 416 HCTNPRPKYREGDVTLPALNLSNVTOSLOLPKOLWSKSVDOYLLPHGKDSILSREVQL 475
DB 420 HCTNTDNPYKREGDLTLAYNLHNVTYLRPLPYFSPNKKQVDKYLRLPLGPHGLLSKSVQL 479
QY 476 NGRLLQWVDDETLPALHEMALAPGSTLGLPAFSYGFVYVIRNAKAIACI 523
DB 480 NGLTLKWVDQDTLPPLMEKPLRPOSSLGLPAFSYGFVYVIRNAKVAACI 527
RESULT 10
ID ADO63826
AD ADO63826 standard; protein; 527 AA.
XX AC ADO63826;
XX DT 26-AUG-2004 (first entry)
XX DE Chimeric heparanase mutant E343A, SEQ ID:11.
XX KW Human; chicken; heparanase; heparanase-derived protein;
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency; drug screening;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome;
KW vulnery; mutant; mutein.
XX OS Homo sapiens.
OS Gallus gallus.
OS Synthetic.
OS Chimeric.
XX Key Location/Qualifiers
FT Peptide 1..18
FT /note= "Chicken heparanase signal peptide"
FT Region 19..527
FT /note= "Corresponds to residues 35-543 of human
FT heparanase mutant E343A (SEQ ID NO:8)"
FT Active-site 209
FT /note= "Active site proton donor"
FT Misc-difference 327
FT /note= "Ala replaces wild-type Glu (active site
FT nucleophile). Corresponds to residue 343 of human
FT heparanase mutant E343A (SEQ ID NO:8)"
XX WO2004048558-A2.
XX 10-JUN-2004.
XX 24-NOV-2003; 2003WO-IL000989.
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX N-PSDB; ADO63820.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
XX Claim 10; SEQ ID NO 11; 128pp; English.
XX The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a chimeric
CC protein comprising the signal peptide of chicken heparanase and residues
CC 35-543 of the human heparanase mutant E225A.
XX Sequence 527 AA;
SQ

CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be used to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a chimeric
CC protein comprising the signal peptide of chicken heparanase and residues
CC 35-543 of the human heparanase mutant E343A.
XX
SQ Sequence 527 AA;

Query Match 62.7%; Score 1711; DB 8; Length 527;
Best Local Similarity 62.9%; Pred. No. 3.7e-164;
Matches 332; Conservative 83; Mismatches 107; Indels 6; Gaps 3;
QY 1 MLVLLLVLLVPPRTA---ELQGLREPIGAVSPFLSLTLDASLARDPRFVALLR 56
DB 1 MLVLLLVLLVPPRTAQDVVDLDFTEQLPLHLVSPFLSLTLDANLATDPRFLILG 60
QY 57 HPKHLTLASGLSGFLRFGTSTDFLFPNKDSTWEEKVLSFQA-KDVCFAWPSFAVY 115
DB 61 SPKLRLTARGLSPAYLRFPGTKTDFLIDPKKSTFPEERSYFWSQVNDICKYSGIPDV 120
QY 116 PKLLLTQWPLQEKLLAEHKKHNTTITRSTDLIDLTFTASSGFLRVFLGNALLRRAG 175
DB 121 EEKRLWEPQEQQLLREHVQKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLTAD 180
QY 176 LQWSSNAKQLLYCAQRSYNIWELGNPNRFRKSGICIDGFLQGRDFVHLRLQSLQH 235
DB 181 LQWSSNAQLLDYSSKGYNIWELGNPNRFLKADIFINGSQLGEDFTQLHKL-LRK 239
QY 236 PLRYHAELGLDVGQPKRKHQHLRLSPMKSGGKAIDSVTHHHYVNGRSATREDFLSPV 295
DB 240 STEFNAKLYGPDVGQPRRKAKMLKSLFKAGGEVIDSVTHHHYVNGRTATREDFLNPV 299
QY 296 LDSFATAIHDLVGLVEATVPKVKVLTGETSGAYGGAPQLSNTYVAGFMWLDKGLAARR 355
DB 300 LDIFISVQKVFQVVESTREPKVWLGTSAVGGAPLSDTPAAGFMWMDKGLGARM 359
QY 356 GIDVWRQVSFGAGSYHLVDAGFKPLDPDYWLSSLYKRLVGRVTLQASVEQADARRPVYL 415
DB 360 GIEVWRQVFFGAGNYHLVDENPDYWLSSLFKLVGTVKLVMSVQSGKRLRVYL 419
QY 416 HCTNPRHPKYREGDVTFLALNLSNVTQSLQPKQWMSKSDVQYLLPHGKDSILSRVQL 475
DB 420 HCTNTDNPRYKEGDTLYALNHNVTKYRLPLYPFNSKQVDKYLLRPLGPHGLLSKSVQL 479
QY 476 NGRLLQWDDTETPALHEMALPGSTLGLPAFSYGFVIRNAKAIACI 523
DB 480 NGLTLKWDDQTLPLMEKFLRPGSSGLPFAFSYFFVIRNAKVAACI 527

RESULT 11
ADO63827
ID ADO63827 standard; protein; 527 AA.
XX
AC ADO63827;
XX

DT 26-AUG-2004 (first entry)
XX Chimeric heparanase mutant E225A/E343A, SEQ ID:12.
DE
XX Human; chicken; heparanase; heparanase-derived protein;
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnery; mutant; mutein.
XX
OS Homo sapiens.
OS Gallus gallus.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /note= "Chicken heparanase signal peptide"
FT Region 19..527
FT /note= "Corresponds to residues 35-543 of human
FT heparanase mutant E225A/E343A (SEQ ID NO:9)"
FT Misc-difference 209
FT /note= "Ala replaces wild-type Glu (active site proton
FT donor). Corresponds to residue 225 of human heparanase
FT mutant E225A/E343A (SEQ ID NO:9)"
FT Misc-difference 327
FT /note= "Ala replaces wild-type Glu (active site
FT nucleophile). Corresponds to residue 343 of human
FT heparanase mutant E225A/E343A (SEQ ID NO:9)"
PN WO2004048558-A2.
XX
PD 10-JUN-2004.
XX
PF 24-NOV-2003; 2003WO-IL000989.
XX
PR 24-NOV-2002; 2002IL-00153059.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
DR N-PSDB; ADO63821.
XX
XX New nucleic acid construct comprising heparanase-derived polypeptide,
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
PT thrombasthenia, or Bernard-Soulier syndrome.
PS
PS Claim 10; SEQ ID NO 12; 128pp; English.
CC
CC The invention relates to nucleic acid constructs comprising a nucleic
CC acid encoding a heparanase-derived protein which lacks heparanase
CC endoglycosidase catalytic activity but which retains its cell-cell and
CC cell-matrix adhesion properties. The constructs of the invention
CC optionally further comprise operably linked regulatory elements. The
CC invention also relates to the heparanase-derived proteins and host cells
CC comprising the nucleic acid constructs of the invention. The heparanase-
CC derived proteins are especially mutants of human heparanase in which the
CC active site proton donor Glu225 and/or the active site nucleophile Glu343
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may
CC optionally further comprise an avian heparanase signal peptide (ADO63825-
CC ADO63827). The heparanase-derived protein, nucleic acid construct and
CC host cells are useful in preparing a tissue sealant composition for
CC sealing injuries, reducing the loss of blood, accelerating the healing
CC and homeostasis of an injury, accelerating blood vessel endothelium
CC formation or the endothelialisation of vascular grafts, accelerating the
CC adhesive activity of mammalian cells, and accelerating the adhesion and
CC aggregation of platelets. They may also be used in the treatment of
CC disorders associated with adhesion deficiency such as LAD (leukocyte
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet

CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
CC cells of the invention may additionally be to screen for modulators of
CC cell-cell and cell-matrix adhesion, and to prepare an implantable
CC synthetic vascular graft comprising a tube made of a biocompatible
CC material lined with the cells. The present sequence represents a chimeric
CC protein comprising the signal peptide of chicken heparanase and residues
CC 35-543 of the human heparanase double mutant E225A/E343A.

XX Sequence 527 AA;

Query Match 62.5%; Score 1705; DB 8; Length 527;
Best Local Similarity 62.7%; Pred. No. 1.5e-163;
Matches 331; Conservative 83; Mismatches 108; Indels 6; Gaps 3;

QY 1 MLVLLLVLLAVPPRTA---ELQGLREPIGAVSPAFSLTLTLDASLARDPRFVALLR 56
DB 1 MLVLLLVLLAVPPRTAQDVVDLDFEFTQEPHLVSPFSLVTIDANLATDRFLILG 60
QY 57 HPKLHTLASGLSPGLRFGTSTDFLIFPNKOSTWEEKVLSFQA-KDVCEAWPSPAVV 115
DB 61 SPKRLTLARGSLPAYLRFGGTKTDFLIDPKKESTFEERSYQSQVNDICKYGSIPDV 120
QY 116 PKLLLTQWPLQEKLLAEHSWKHKNTTITRSTLDILHTFASSGFLVGLNALLRRAG 175
DB 121 EKKRLDWPYQEQLLREHYQKKPKNSTYSRSVDVLYTFANCSDGLDIFGLNALURTA 180
QY 176 LOWDSSNAQLCYCAQRSYNISWELNEPNPFRKSGICIDGFLQGRDFVHLRQLLSQH 235
DB 181 LQWSSNAQLLDYCSKGNISWELGNAPNPFKXADIFNGSQLGEDFIQLHKL-RK 239
QY 236 PLYRHAELGDLVGQPRKTHQLLRSPFMKSGGKAIDSVTWHYYVNGRSATREDFLSPEV 295
DB 240 STFKNAKLVGPDVGQPRKTKAKMLKGLKAGGEVIDSVTHYYVNGRTATREDFLNPDV 299
QY 296 LQSFATAIHDLGIVATVPCKVWLGTSATGAGGAPQLSNYVAGFMWLDKGLAARR 355
DB 300 LDIFISVQVQFQVSTREKGVWLGATSSAYGGGAPLLSDTFAAGFMWLDKGLSARM 359
QY 356 GDVVMRQVSFGAGSVHLVDAGPKPLPDYWLILLYKRLVGTTRVLOASVEQADARRPRVYL 415
DB 360 GIEVVMRQVFFGAGNVHLVDENFDPLDYWLILLYKRLVGTTRVLOASVEQADARRPRVYL 419
QY 416 HCTNPRHPKYREGDVTFLFALNLSNVTQSLQPLKQLMSKSDVQYLLPLPHGKDSILSREVQL 475
DB 420 HCTNTDNPYKEGDLTYALNHNVTYLRPLPYFSPKQVDKYLLRPLGPHGLLSKSVQL 479
QY 476 NGRLLQWDDTETLPAHEMALAGSTGLPAPSYGFVIVIRNAKAIACI 523
DB 480 NGLTLKWDQDTLPLMEKPLRPGSSILGLPAFSYGFVIRNAKVAACI 527

RESULT 12

ID ADO63832

XX ADO63832 standard; protein; 543 AA.

AC ADO63832;

XX 26-AUG-2004 (first entry)

DE Human heparanase mutant E396A.

XX Human; heparanase; heparanase-derived protein; heparanase mutant;
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
KW vascular graft; platelet adhesion; platelet aggregation;
KW adhesion disorder; LAD; leukocyte adhesion deficiency;
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
KW vulnarery; mutant; muten; enzyme.

XX Homo sapiens.

OS Synthetic.

XX key Location/Qualifiers

FT Active-site 225 /note= "Active site proton donor"
FT Active-site 343 /note= "Active site nucleophile"
FT Misc-difference 396 /note= "Ala replaces wild-type Glu"
XX WO2004048558-A2.
XX 10-JUN-2004.
XX 24-NOV-2003; 2003WO-IL000989.
XX 24-NOV-2002; 2002IL-00153059.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;
XX WPI; 2004-450373/42.
XX New nucleic acid construct comprising heparanase-derived polypeptide,
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's
XX thrombasthenia, or Bernard-Soulier syndrome.
XX Example 4; Page; 128pp; English.
XX The invention relates to nucleic acid constructs comprising a nucleic
XX acid encoding a heparanase-derived protein which lacks heparanase
XX endoglycosidase catalytic activity but which retains its cell-cell and
XX cell-matrix adhesion properties. The constructs of the invention
XX optionally further comprise operably linked regulatory elements. The
XX invention also relates to the heparanase-derived proteins and host cells
XX comprising the nucleic acid constructs of the invention. The heparanase-
XX derived proteins are especially mutants of human heparanase in which the
XX active site proton donor Glu225 and/or the active site nucleophile Glu343
XX are replaced with Ala (ADO63822-ADO63824), and the proteins may
XX optionally further comprise an avian heparanase signal peptide (ADO63825-
XX ADO63827). The heparanase-derived protein, nucleic acid construct and
XX host cells are useful in preparing a tissue sealant composition for
XX sealing injuries, reducing the loss of blood, accelerating the healing
XX and homeostasis of an injury, accelerating blood vessel endothelium
XX formation or the endothelialisation of vascular grafts, accelerating the
XX adhesive activity of mammalian cells, and accelerating the adhesion and
XX aggregation of platelets. They may also be used in the treatment of
XX disorders associated with adhesion deficiency such as LAD (leukocyte
XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet
XX function), or Bernard-Soulier syndrome (deficient platelet adhesion). The
XX cells of the invention may additionally be to screen for modulators of
XX cell-cell and cell-matrix adhesion, and to prepare an implantable
XX synthetic vascular graft comprising a tube made of a biocompatible
XX material lined with the cells. The present sequence represents a human
XX heparanase mutant E378A created in an example of the invention which
XX retains its heparanase catalytic activity. The present sequence is not
XX shown in the invention, but is derived from the protein sequence of
XX GenBank accession number AF144325 and the information provided on page
XX 70.

SQ Sequence 543 AA;

Query Match 60.6%; Score 1653.5; DB 8; Length 543;
Best Local Similarity 60.5%; Pred. No. 2.7e-158;
Matches 322; Conservative 86; Mismatches 113; Indels 11; Gaps 3;

QY 2 LVLLLLLVLLAVPP-----RRTAELQGLREPIGAVSPAFSLTLTLDASLARDPRFV 52
DB 13 LMLLLGLPLSPGALPRPAQADVVLDLDFEFTQEPHLVSPFSLVTIDANLATDRPL 72
QY 53 ALLRHPKLLHTLASGLSPGLRFGTSTDFLIFPNKOSTWEEKVLSFQA-KDVCEAWPSP 111
DB 73 ILGSPKRLTLARGSLPAYLRFGGTKTDFLIDPKKESTFEERSYQSQVNDICKYGS 132
QY 112 FAVVPKLLTQWPLQEKLLAEHSWKHKNTTITRSTLDILHTFASSGFLVGLNALL 171

Db 133 PPVVEKRLLEWYQEQQLLREHYQKFKNSTYSSVDVLYTFANCSGLDLIFGLNALL 192
Qy 172 RRAGLQWSSNAQQLGYCAQRSYNIWELGNENPNSFRKSGICIDGFLGRDFVHLRQL 231
Db 193 RTADLQWSSNAQQLLDYCSKGYNIWELGNENPNSFLKKADIFINGSQLGEDFIQLHKL 252
Qy 232 LSQHLRYRHAELVGLDVGPQRKHTQHLRLRFMSKGGKKAIDSVTWHYVNGRSATREDFL 291
Db 253 L-RKSTFKNAKLYGPDVGPQRKHTAKMLKSLFKAGGEVIDSVTWHYVNGRSATREDFL 311
Qy 292 SPEVLDSFATAIHDLVIGVEATVPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKLGL 351
Db 312 NPVDLIDIFISSVQKVFQVVESTRPGKVMVGETGSAYGCGAPLSDTFAAGFWMLDKLGL 371
Qy 352 AARRGIDVVMRQVSFGAGSHLVDAFGKPLPDYWLSSLYKRLVGRVLRQASVEQADARRP 411
Db 372 SARMGIEVVMRQVFFGAGNYHLVDANFDPLPDYWLSSLYKRLVGRVLRQASVEQADARRP 431
Qy 412 RYVLHCTNPRHPKYRGDVTLPALNLSNVTQSLQPKQLMSKSDVQVLLPLPHGKDSILSR 471
Db 432 RYVLHCTNTPRYKEGDLTYAINLHNVTYKRLPYFPFNKQVYDKYLLRPLGPHGLLSK 491
Qy 472 EVOLNGLRLQWDDTLPALHEMALAPGSTLGLPAESYGFVIRNAKAIACI 523
Db 492 SVQNLGLTLKQWDDQTLPLMEKPLRPGSSGLPFAFSYSPFVIRNAKAAACI 543

RESULT 13

ABP56822
ID ABP56822 standard; protein; 545 AA.
XX
AC ABP56822;
DT 02-APR-2003 (first entry)
XX
DE Human heparanase protein SEQ ID NO:18.
XX
KW Human; heparanase; phosphorothioate; antisense oligonucleotide;
KW cytotatic; gene therapy; tumour; enzyme.
OS Homo sapiens.
XX
XX WO2003004705-A1.
XX
XX 16-JAN-2003.
XX
XX 01-JUL-2002; 2002WO-US020636.
XX
XX 05-JUL-2001; 2001US-00899440.
XX
XX (UYCO) UNIV COLUMBIA NEW YORK.
XX

Stein C;

XX
XX WPI; 2003-201558/19.
XX N-PSDB; AB222816.
XX
XX New oligonucleotide having a sequence complementary to a sequence of
XX ribonucleic acid encoding a heparanase, useful for preparing a
XX composition for treating tumor.
XX
XX Disclosure; Page 46-47; 48pp; English.
XX
XX The present invention describes an oligonucleotide having a sequence
XX complementary to a sequence of ribonucleic acid encoding a heparanase.
XX The oligonucleotide hybridises with the ribonucleic acid under conditions
XX of high stringency and has a sequence comprising 10-40 bp. The
XX internucleoside linkages of the oligonucleotide comprise at least one
XX phosphorothioate linkage. Hybridisation of the oligonucleotide to the
XX ribonucleic acid inhibits expression of the heparanase, where inhibition
XX of heparanase means at least a 50% reduction in the quality of
XX heparanase. Also described: (1) a method of inhibiting expression of a

CC heparanase in a cell; (2) a composition comprising the above
CC oligonucleotide in an amount effective to inhibit the expression of
CC heparanase in the cell and a carrier; and (3) a method of treating a
CC tumour in a subject comprises administering to the subject an amount of
CC the above oligonucleotide effective to inhibit expression of a heparanase
CC in the subject. Heparanase antisense oligonucleotides have cytostatic
CC activity, can be used in gene therapy, and can be used for preparing a
CC composition for treating tumours. The present sequence represents human
CC heparanase, which is given in the exemplification of the present
XX invention

SQ Sequence 545 AA;

Query Match 60.5%; Score 1650; DB 6; Length 545;
Best Local Similarity 60.1%; Pred. No. 6.3e-158;
Matches 321; Conservative 87; Mismatches 114; Indels 12; Gaps 3;

Qy 1 MLVLLLVLLAVPP-----RRTAELQLGREPIGAVSPAPLSLTLTDLASLARDPR 50
Db 13 LMLLLGLPLGSPGALPRPAQAQDDVDDFFQEPHLHLVSPFLSVITDANLATDPR 72
Qy 51 FVALLRHPKLTSLASGLSPGFLRFEGGTSTDLIFNPKNKDSWEEKVLSFQA-KDYCEAM 109
Db 73 FLILLGSPKLTTLARGLSPAVLRFGGTTDFLFPKXSTFEERSYMQSVQVNDICKYG 132
Qy 110 PSFAVVPKLLTQWPLQEKLLLAHSHWKHKNTTITRSTLDLHTFASSSGRPLVGLNA 169
Db 133 SIPPVVEKRLLEWYQEQQLLREHYQKFKNSTYSSVDVLYTFANCSGLDLIFGLNA 192
Qy 170 LLREAGLQWSSNAQQLGYCAQRSYNIWELGNENPNSFRKSGICIDGFLGRDFVHLR 229
Db 193 LLRTADLQWSSNAQQLLDYCSKGYNIWELGNENPNSFLKKADIFINGSQLGEDFIQLH 252
Qy 230 QLLSQHPLYRHAELVGLDVGPQRKHTQHLRLRFMSKGGKKAIDSVTWHYVNGRSATRED 289
Db 253 KLL-RKSTFKNAKLYGPDVGPQRKHTAKMLKSLFKAGGEVIDSVTWHYVNGRSATRED 311
Qy 290 FLSPVLDSFATAIHDLVIGVEATVPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKL 349
Db 312 FLNPDVLDIFISSVQKVFQVVESTRPGKVMVGETGSAYGCGAPLSDTFAAGFWMLDKL 371
Qy 350 GLAARRGIDVVMRQVSFGAGSHLVDAFGKPLPDYWLSSLYKRLVGRVLRQASVEQADAR 409
Db 372 GLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLYKRLVGRVLRQASVEQADAR 431
Qy 410 RPRVYLHCTNPRHPKYRGDVTLPALNLSNVTQSLQPKQLMSKSDVQVLLPLPHGKDSIL 469
Db 432 KLRVYLHCTNTPRYKEGDLTYAINLHNVTYKRLPYFPFNKQVYDKYLLRPLGPHGLL 491
Qy 470 SREVQLNGLRLQWDDTLPALHEMALAPGSTLGLPAFSYGFVIRNAKAIACI 523
Db 492 SKSVQLNGLTLKQWDDQTLPLMEKPLRPGSSGLPFAFSYSPFVIRNAKAAACI 545

RESULT 14

ADE16012
ID ADE16012 standard; protein; 545 AA.

XX AC ADE16012;

XX DT 29-JAN-2004 (first entry)

XX G-coupled protein receptor related polypeptide, SEQ ID No 42.

XX G-coupled protein receptor; antidiabetic; anorectic; antibacterial;
KW virucide; fungicide; cytostatic; nootropic; neuroprotective;
KW antiparkinsonian; haemostatic; antilipemic; neurogenesis;
KW cell differentiation; cell proliferation; hematopoiesis; wound healing;
KW angiogenesis; gene therapy; chromosome mapping; tissue typing;
KW preventive medicine; pharmacogenomics; human.

XX Homo sapiens.

XX

PN WO200283841-A2.
 XX 24-OCT-2002.
 XX 03-APR-2002; 2002WO-US010713.
 XX 03-APR-2001; 2001US-0281136P.
 PR 05-APR-2001; 2001US-0281863P.
 PR 05-APR-2001; 2001US-0281906P.
 PR 10-APR-2001; 2001US-0282934P.
 PR 13-APR-2001; 2001US-0283657P.
 PR 13-APR-2001; 2001US-0283678P.
 PR 13-APR-2001; 2001US-0283687P.
 PR 13-APR-2001; 2001US-0283710P.
 PR 17-APR-2001; 2001US-0284234P.
 PR 19-APR-2001; 2001US-0285325P.
 PR 20-APR-2001; 2001US-0285609P.
 PR 23-APR-2001; 2001US-0285748P.
 PR 23-APR-2001; 2001US-0285890P.
 PR 24-APR-2001; 2001US-0286068P.
 PR 27-APR-2001; 2001US-0287213P.
 PR 03-MAY-2001; 2001US-0288509P.
 PR 30-MAY-2001; 2001US-0294495P.
 PR 31-MAY-2001; 2001US-0294801P.
 PR 31-JUL-2001; 2001US-0309216P.
 PR 25-SEP-2001; 2001US-0324775P.
 PR 28-NOV-2001; 2001US-0333900P.
 PR 02-APR-2002; 2002US-00115479.
 XX (CURA-) CURAGEN CORP.
 XX Li L, Gerlach V, Liu X, Miller CE, Spytek KA, Zerhusen BD;
 PI Pena CE, Shenoy SG, Zhong H, Smithson G, Caaman SJ, Boldog FL;
 PI Voss EZ, Vernet CAM, Macdougall JR, Rastelli L, Anderson DW;
 PI Zhong M, Mezes PD, Furtak K, Patturajan M, Burgess CE, Malyankar UM;
 PI Shinkets RA, Taupier RJ, Edinger SR, Mazur A;
 DR WPI; 2003-067574/06.
 DR N-PSDB; ADL61011.
 XX New isolated NOVX polypeptides and polynucleotides, useful for
 PT preventing, diagnosing or treating NOVX-associated disorders e.g.
 PT diabetes, obesity, dyslipidemias, cancer, Parkinson's disease,
 PT Alzheimer's disease, infections.
 XX Claim 1; SEQ ID NO 42; 320pp; English.

CC represents one of the novel G-coupled protein receptor related
 CC polypeptides of the invention.
 XX
 SQ Sequence 545 AA;
 Query Match 60.5%; Score 1650; DB 7; Length 545;
 Best Local Similarity 60.1%; Pred. No. 6.3e-158;
 Matches 321; Conservative 87; Mismatches 114; Indels 12; Gaps 3;
 QY 1 MLVLLLVLLAVPP-----RRTAELQLGLREPIGAVSPAFSLTLTASLARDP 50
 DB 13 LLMLLLLGPIPLSPGALPRPAQAQDVLDFTFQEPHLVSPFSLVTIDANLATDP 72
 QY 51 FVALLRHPKLTLAGSLSPGFLRFGGTSTDFLI FPNKNDSTWEEKVSEFOA-KDVCEAW 109
 DB 73 FLILGSPKLTTLARGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQSQVNQDICKYV 132
 QY 110 PSFAVVPKLLLTQWPLQEKLLLAHSHWKKHNTTITRSTLILTFPASSSSFLVFGUNA 169
 DB 133 SIPPDVEEKLRLEWPYOEQLLREHYQKFKNSTYSRSSVDVLYTFANCSGLDILFGLNA 192
 QY 170 LLRRAGLQWSSNAQLLGYCAORSYNI SWELGNEPNSFKKSGICIDGFGOLGRDFVHLR 229
 DB 193 LLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKADIFINGSOLGEDFIQLH 252
 QY 230 QLLSQHPLYRHAELYGLDVGPQRKHTQHLLRSFMKSGKKAIDSVTWHYYVNGRSATRED 289
 DB 253 KLL-RKSTFNKLYGPDVGPQRKTA KMLKFLKAGEVIDSVTWHYYLNGRTATRED 311
 QY 290 FLSPEVLDSFATAIHDVLGIVEATVPCKKWLGTSGSAYGGAPOLSNYYVAGFWMLDKL 349
 DB 312 FLNPDVLIDFISSVQKVFQVVESTRPCKKWLGTSGSAYGGAPLLSDTFAAGFWMLDKL 371
 QY 350 GLAARGIDVVMQVSGAGSYHLVDAGFKPLDPYWLSSLYKRLVGFVLOASVEQADAR 409
 DB 372 GLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTKVLMASVQGSRRR 431
 QY 410 RPRVYLCTNPRHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVDQTLPLPHGKDSIL 469
 DB 432 KLRVYLCTNPDNPRYKEGDLTYAINLHNVTYLRPLPYPSNKQVDXYLLRPLPHGGLL 491
 QY 470 SREVQLNGRLLQWDDTLPALHEMALAPGSTLGLPAPSYGYFVIRNAKAIACI 523
 DB 492 SKSVQLNGLTLKWVDDQTLPLMEKPLRPGSSILGLPAPSYSFYFVIRNAKVAACI 545
 RESULT 15
 ID ADL93951 standard; protein; 545 AA.
 XX AC ADL93951;
 XX DT 20-MAY-2004 (first entry)
 XX DE Human G-coupled protein receptor-related protein #21.
 XX KW human; transgenic; Gene Therapy; Protein Therapy; cardiomyopathy;
 KW atherosclerosis; hypertension; congenital heart defect; aortic stenosis;
 KW atrial septal defect; atrioventricular canal defect; ductus arteriosus;
 KW pulmonary stenosis; subaortic stenosis; ventricular septal defect;
 KW valve disease; tuberosclerosis; scleroderma; obesity; transplantation;
 KW adrenoleukodystrophy; congenital adrenal hyperplasia; prostate cancer;
 KW neoplasia; adenocarcinoma; lymphoma; uterus cancer; fertility;
 KW haemophilia; hypercoagulation; idiopathic thrombocytopenic purpura;
 KW immunodeficiency; graft versus host disease; AIDS; bronchial asthma;
 KW Crohn's disease; G-coupled protein receptor; metabolic disorder;
 KW neurodegenerative disorder; receptor.
 XX OS Homo sapiens.
 XX PN US2004006205-A1.
 XX PD 08-JAN-2004.

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.4126 Seconds
(without alignments)
2889.939 Million cell updates/sec

Title: US-10-645-659A-5
Perfect score: 2728
Sequence: 1 MLVLLLVLLAVPPRTAE.....LPAPSYGYVIRNAKAIACI 523

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: Piri:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	810.5	29.7	480	JC7506	heparanase protein
2	395	14.5	521	T45608	hypothetical prote
3	128.5	4.7	190	T01953	hypothetical prote
4	108.5	4.0	1004	S53939	probable membrane
5	106.5	3.9	1225	C84330	hypothetical prote
6	106	3.9	503	A48546	genome polyprotein
7	105	3.8	473	S71934	genome polyprotein
8	105	3.8	496	GNX31E	genome polyprotein
9	104.5	3.8	374	B82488	hypothetical prote
10	102.5	3.8	1180	AI1939	two-component hybr
11	101	3.7	1076	A69409	carbamoyl-phosphat
12	99	3.6	987	D97029	ribonucleotide red
13	98.5	3.6	836	D64988	yejO protein - Esc
14	97.5	3.6	284	JS0652	aminoglycoside N3'
15	97.5	3.6	780	S44560	alpha, alpha-trehal
16	97.5	3.6	836	A98014	hypothetical prote
17	97.5	3.6	836	C85858	hypothetical prote
18	97	3.6	341	A57136	class I histocompa
19	97	3.6	493	JC1327	protective antigen
20	96.5	3.5	565	A10479	probable membrane
21	96.5	3.5	627	C98148	acetoin catabolism
22	96.5	3.5	627	AI1319	transcription regu
23	95.5	3.5	1012	GNXSAU	genome polyprotein
24	95	3.5	4568	T08030	dyein beta heavy
25	94.5	3.5	395	C84138	8-amino-7-oxononan
26	94	3.4	584	AH2506	ser/thr protein ki
27	94	3.4	867	D87320	conserved hypothet
28	94	3.4	897	G02529	dyein heavy chain
29	93.5	3.4	485	T07596	1-aminocyclopropan

30	93.5	3.4	492	2	F70326	conserved hypothet
31	93.5	3.4	500	2	D87541	beta-xylosidase [i
32	93.5	3.4	567	2	T44363	poly(3-hydroxybuty
33	93	3.4	401	2	B83377	probable phospholi
34	93	3.4	515	2	AE0186	probable decarboxy
35	93	3.4	1039	2	C87083	C-term lysyl-tRNA
36	92.5	3.4	341	2	JC5663	major histocompati
37	92	3.4	893	2	T15183	hypothetical prote
38	92	3.4	1839	1	OYBYK	adenylate cyclase
39	92	3.4	2108	2	H70819	probable polyketid
40	91.5	3.4	454	2	JC4616	apyrase (EC 3.6.1.
41	91.5	3.4	964	2	S45944	hypothetical prote
42	91.5	3.4	1012	1	GNXS52	genome polyprotein
43	91.5	3.4	1379	2	T13718	pollux gene protei
44	91	3.3	884	2	H83322	hypothetical prote
45	91	3.3	911	2	A56465	transcription fact

ALIGNMENTS

RESULT 1

JC7506
heparanase protein 2a - human
C:Species: Homo sapiens (man)
C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C:Accession: JC7506
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me
A:Reference number: JC7506
A:Accession: JC7506
A:Molecule type: mRNA
A:Residues: 1-480 <MCK>
A:Cross-references: UNIPROT:Q9HE39; UNIPARC:UPI000003E88A; GB:AF282885
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and therai
therapies.
C:Genetics:
A:Gene: hpa2a
A:Map position: 10q23-10q24
C:Keywords: heparin binding; membrane bound

Query Match	29.7%	Score	810.5	DB	2	Length	480
Best Local Similarity	34.8%	Pred. No.	5.1e-58				
Matches	192	Conservative	69	Mismatches	161	Indels	129
Gaps	9						
Qy	1	MLVLLLVLLA-----VPPRTAEIQ-----LGLREPICAVSPAFSLTLTLDAS	44				
Db	27	LYLALLHLSSQNGDRRPIVDRAAGLKEKTLILLDVTSTKNPVTNVNFIQLQDPS	86				
Qy	45	LARDPRFVALLRHPKLTLAGSLSPGLRFGTSTDFLIF-----NPNKOSTWEEKVLSEF	100				
Db	87	IIHD-GWLDFLSSKRLVTLARGLSPAFLRFGKRTDFLQFQNLNPAKS-----	134				
Qy	101	OAKDVCEAWPFAVVPKLLLTQWPLOEKLLLAHSHKXKNTTITRSLDILTTFASSSG	160				
Db	135	-----	134				
Qy	161	FRLVFGLNALLRRAGLQWDSNAKQLLGYCAQRSYNISWELGNEPNSFRKSGICIDGFG	220				
Db	135	-----RGFGPD-----YYLKNYE--DEPNNTYTMGRAVNGSQ	166				
Qy	221	LGRDFVHLRQLLSOHLRYHAEYGLDVGQPKRHTQHLLRSFPMKSGKAIDSVTWHYYV	280				
Db	167	LGDYIQLKSLQPIRIYRSRSLYGNIGRPRKNVIALLDGPMKVAGSTVDATVWQHCVI	226				
Qy	281	NGRSATREDFLSPVLDLSFAITHDVLGIVATVPKKVWLGETSGAYCGGAPQLSNTYV	340				
Db	227	DGRVVKVMDFLKTRLLDLSQIRKIQKVNNTYTPGKKIWLGVVTTTSAGGTNNLSDSA	286				
Qy	341	AGFWLWDLGLAARGIDVVMRQVSGAGSHLVNDAGFKPLPDYWLISLLYKRLVGRVLQ	400				
Db	287	AGFLWLTGLMANOGIDVIRHSFFDHGYNHLVDQNFNPLPDYWLISLLYKRLIGPKVLA	346				

```
QY 401 ASVEQADAR-RP-----RVYLHCTNPRHPKYREGDVTFLFALNLSNVTQSLQPKQLW 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 VHVAGLQRPGRVIRDKLRIYAHCTNHNHNHYVRGSIITFLIINLHRSRKIKLAGTLR 406

QY 452 SKSVQVYLLPHGKOSILREVQLNGRLQVDDDTLPALHEMALAPGSLGLPARSYGF 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 DKLVHQYLLQPYQEGELKSKSVQLNGOPLVMVDDGTLPKLPRLRAGRTRLVIPPVTMGF 466

QY 512 YVIRNAKAIAIC 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 467 FVVKVNNALAC 477

RESULT 2
T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UPI000000A497C; EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A:Note: F13G24.30

Query Match 14.5%; Score 395; DB 2; Length 521;
Best Local Similarity 26.2%; Pred. No. 4.3e-24;
Matches 137; Conservative 79; Mismatches 195; Indels 112; Gaps 19;

QY 55 LRHPKLTLASGLSPGFLRPGTSTDFLPNPKDSTWEKVLSEFQAKDVCEAWPSFAV 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 LTRPLLTAKAKAPKPLRIIRIGGSLQDQVIYDVG-----NLKTPCR----- 94

QY 115 VPKLILQTWPLOQ-----KLLLAHSWKKHKNTTITRSTLDILHTFASSGFLRVF 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 -----PFQWNSGLFGFSKGCLMKKW-----DELNSPLTATGAVTF 132

QY 166 GLNALRRAGLQ-----WDSSNAKOLLGYCAQRSYNI-SWELGNEPNSFRKKSGLCID 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 GLNALGRHKLKRGKAGGAWDHINTQDFLNTYTSKGVIDSWEFGNELSG--SGVGASVS 190

QY 218 GFQLGRDFVHLROLLSQHPLYRHAELY-GLDVQGPQRKHTOHLRSFMKSGKKAIDSVTWH 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 AELYGKDLIVLKDVIK--VYKNSWLHKPLVAPGGFYEQQWYTKLLEISGPSVDVVTTH 248

QY 277 HYVNGRS---ATREFLSPEVLDSFATAIHDVLGIVEATVPKKVWLGTGSAYGCGAP 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 HYNLGSNDPALVKIMDPSYLSQVSKTFKDVNQTIQEHGPWASPMWGESGGAYNSGGR 308

QY 334 QLSNTYVAGFMWLDKGLAARRGIDVVMRQVSPFAGSYHLVDAG-FKPLPDYMLSLLYKR 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 HVSDTFIDSPWYLDQLGMSARHNTKYVCRQTLVG-GFYGLEKGTFPVNPEDYYSALLWHR 367

QY 393 LVGTRVLQASVEQADARRPRVYLHCTNPRHPKYREGDVTFLFALNLSN----- 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 LMGKGVL--AVQTDGPPQLRVAHCSKGR-----AGVTLLILNLSNQSDFTVSVNSGIN 419

QY 440 -----VTQSIQLP-KOLWSKSDVQYL-----LLPHGKDSIL-SREVQLNGRL 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 VVLNBSRRKKSLDITLKRFPFSGWIGSKASDGYLNREYHLP--ENGVLRSKTMVLNGKS 477

QY 480 LQWVDDETLPALHEMALAPGSLGLPARSYGFYVIRNAKAIAIC 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 LKPTATGDIPLSEPLVRLSVNSPLNVLPLSMFVLPNFDASAC 520
```

RESULT 3

```
T01953
hypothetical protein T2L5.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C:Accession: T01953
R:Geisel, C.; Smith, A.; Le, T.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana T2L5.
A:Reference number: Z14470
A:Accession: T01953
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-190 <GEI>
A:Cross-references: UNIPROT:O82604; UNIPARC:UPI000000A8F7D; EMBL:AF096371; NID:g3695386; I
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 36/2; 69/3
A:Note: T2L5.6
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 4.7%; Score 128.5; DB 2; Length 190;
Best Local Similarity 25.1%; Pred. No. 0.0052;
Matches 47; Conservative 35; Mismatches 70; Indels 35; Gaps 8;

QY 362 RQVSPGAGSHLYVD-AGFKPLPDYWLSSLYKRLVGTIVLQASVEQADARRPRVYLHCTNP 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 RQSLIG-GNYGLLNTTNFTNPDPYSALIWRLMGRKALFTFP--SGTKKIRSYTHCAR- 67

QY 421 RHPKYREGDVTFLFALNLSNVTQ-----SLQLPKOL--WSKSDVQVLLPLPHG--- 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 -----QSKGITVLLMLNDNTTIVVAKVELNNSFSLRTHKMKSYKRASSQLFGPGNGVIQ 122

QY 465 -----KDSIL-SREVQLNGRLQVDDDTLPALHEMALAPGSLGLPARSYGFYVIR 515
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 REEVHLTAKDGNLHSQTMLLNGNALQVNSMGDLFPPIPIHINSTEPIPIAFYSIVFVHMR 182

QY 516 NAKAIAIC 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 NVVVPAC 189

RESULT 4
S53939
probable membrane protein YGL241w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G0733; hypothetical protein HRC1004
C:Species: Saccharomyces cerevisiae
C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S53939; S60489; S64266
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, April 1995
A:Description: The sequence of a 11.1 kb DNA fragment between ADH4 and ADE5 on the left e
A:Reference number: S53934
A:Accession: S53939
A:Molecule type: DNA
A:Residues: 1-1004 <VAN>
A:Cross-references: UNIPROT:P53067; UNIPARC:UPI0000052FA2; EMBL:Z49149; NID:g793865; PID:
A:Experimental source: strain S288C
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
Yeast 11, 1519-1523, 1995
A:Title: The sequence of an 11.1 kb DNA fragment between ADH4 and ADE5 on the left arm of
A:Reference number: S60484; MUID:96353434; PMID:8750240
A:Accession: S60489
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Residues: 1-1004 <VAN>
A:Cross-references: UNIPARC:UPI0000052FA2; EMBL:Z49149; NID:g793865; PIDN:CAA89014.1; PII
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, May 1996
```

A:Reference number: S64263
A:Accession: S64266
A:Molecule type: DNA
A:Residues: 1-1004 <VAF>
A:Cross-references: UNIPARC:UPI0000052FA2; EMBL:Z72763; NID:g1322906; PID:e243899; PID:G
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:KAP114
A:Cross-references: SGD:S0003210
A:Map position: 7L
C:Keywords: transmembrane protein
F:186-202/Domain: transmembrane #status predicted <TMM>

Query Match 4.0%; Score 108.5; DB 2; Length 1004;
Best Local Similarity 20.8%; Pred. No. 2.6;
Matches 120; Conservative 84; Mismatches 212; Indels 161; Gaps 29;

Qy 16 RTAEQL--GLREPIGAVSPAFSLTLDSARDPRFALLRHPKLHTLASGLSPGFRLR 73
Db ||||| : : : : :
Qy 18 REVAETQLLOWCDSDASQVFKAALANALQHESLESQFALLSRKLITMY--WSPGGES 75
Db ||||| : : : : :
Qy 74 FGGTST-----DF-----LIFFPNKDSTWEEKV---LSEFOAKDVCEAWPS-FAVVP 116
Db ||||| : : : : :
Qy 76 YRSTNVEIDVKDFIREVLKLCINDENTYKIKNGASYCIVQISAVDPDQWPOLLTIYI 135
Db ||||| : : : : :
Qy 117 KLLLTQWPLQEKLLEAE----HSWKHKHNTTITRSTDILHTTFASSSGFRLVFGNAL- 170
Db : : : : :
Qy 136 DAISHOHLNSLMSLLNEIYDDVVSEEMFPFGGIGLATMEIVFKVLNTTETSTLIAIAALK 195
Db : : : : :
Qy 171 LRRAGL-----QWDSSNAKQLLYCAQRSYNISWEI-----GNPNSFRKKSICIDG 218
Db : : : : :
Qy 196 LLKACLLQWSSHNEYDEASKSFVSQC LATSLQILGQLLTNLFGN-----VD- 242
Db : : : : :
Qy 219 FOLGRDFVHLROLLSOHPLYRHAELVGLDVGPQRKH-TQHLRSFMKSGGKAIDSVTWHH 277
Db : : : : :
Qy 243 -----VISQLKFKSIYENLVFIKNDFS--RKHFSELQOKFKIMAIQLENVT--- 289
Db : : : : :
Qy 278 YYNGRSATRE-----DFL-----SPEVLDSFATAIHDLVGIVEAT 313
Db : : : : :
Qy 290 -HINAVETTESEPLETWHDCSIYIEFLTVSVCTLOFSVEEMNKIITSLTILCOLSSET 348
Db : : : : :
Qy 314 VPKKWLGETSAYGGGAPQLSNTRYAGFWMLDKGLAARGIDVMVRQVSGFAGSYHL 373
Db : : : : :
Qy 349 ---REIWTS-----FNTEFS-----KETGLAA-----SYNV 372
Db : : : : :
Qy 374 VDAG---FKPLPDYMLSLYLKRVGLGVTRVLOASVEQADARRPRVYLHCNTPRHPPKYREGDV 430
Db | : : : : :
Qy 373 RDQANEFTSLNPQSLIFK-VSNDIEHSTCNYSLESLLYLQCI-----LNDDDEI 426
Db : : : : :
Qy 431 TLFALNLSNVTSQQLPKQLWSKVQDYLLPHGKOSILREVQNGRLLQMWDDETLP A 490
Db : : : : :
Qy 427 T-----GENIDQSL-----QILIKTNLIVSQEIPPELLARAILTIPRVLDKFD-ALPD 476
Db : : : : :
Qy 491 LHEMA---LAPGSTGLP-----AFSYGYF 512
Db : : : : :
Qy 477 IKPLTSAFLAKSLNALKSKDELIKSATLIAFTYYCY 513
Db : : : : :

RESULT 5
C84530
hypothetical protein At2g15540 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84530
R.;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84530
A>Status: preliminary
A:Molecule type: DNA

A;Residues: 1-1225 <STO>
A;Cross-references: UNIPROT:Q9ZQF8; UNIPARC:UPI00000A75F2; GB:AE002093; NID:g4335720; PFI
C;Genetics:
A;Gene: At2g15540
A;Map position: 2

Query Match 3.9%; Score 106.5; DB 2; Length 1225;
Best Local Similarity 20.9%; Pred. No. 5.1;
Matches 126; Conservative 86; Mismatches 233; Indels 159; Gaps 34;

Qy 10 LLAVPPRTAELQLGLREPIG--AVSPAFLSLTLDASLARDPFRVALLRH-----PKLH 61
Db 356 LOEVPVLITEMNKSLLTKVISPVEVKRALFSLNPDKAFGPDGMTAFFQHYQWDLTGPDLL 415
Qy 62 TLASGL-SPGFLREGGTSTDFLFNPKNKSTWEEKVLSEFOAKDYCEAMPSPAVPKLL- 119
Db 416 KLVQNFHSTGTFDERLNETNICLI-----PKTERPKMAEFRPISLCNV--SYKVISKVL 469
Qy 120 -----LTQWPOELKLLAEHSWKKKNTYITRSTLDILTFPASSG 161
Db 470 SRLKRLPELITSETQSFAVERLITDNLIAQENFHALRTNPACCKYKMAINTDMSKAYD 529
Qy 162 RLVPFG-LNALLRRAG-----LOWDSSNAKQLLGYC-AQSSYNIWELGNEPNSFRKSGI 214
Db 530 RVEWSEFRLALMLKMGFAQKWDW-----IIFCISVSYSKIL--LNGSPKGFIRPS-- 577
Qy 215 CIDGFOGLGRDFVHLRQLLSQHPLYR-----HAELYGLDVGPQRKHTOHLIRS-----F 262
Db 578 --RGIROGDPISPLFLICTEALVAKUKDAEWHGRIQGLISRASPSITSHLLFADDSLFF 635
Qy 263 MKSG---GKAIDSTVWHYYNGRSATREDPLSEVLDSFATAIHDVLGIVEATVPKKV 319
Db 636 CKADPLQGEKIIDLRLYGEASGQ-----LNPD--KSSVMFGHEVDNSIRNTI---KV 684
Qy 320 WLGETGSAYGGGA--POLSNYYVAGFMWLDKGLAAREGI-----DVVMRQVSFGAGSYH 372
Db 685 SLG-----IHKDGAIRSKLSSV-VANFWMKTR---EESNGIHWIAWDKLTCTPDSGG----- 732
Qy 373 LVDAGFKPLPDYWLISLYIKRL-----VGTRVLQA-----SVEQADARRPRVYLH 416
Db 733 ---LGFRTEEFNLVLLAKQLWRLIRFPNLSRLVLRGFRYSPDIQIGKANRPSFGWR 789
Qy 417 CTPNRPHPKREGDVTLLFALMLSNVTSGL--OLPKQLMSKSVDDQYL--LLPHGKDSILS-R 471
Db 790 SIMAAKP-----LLLSGLRRTIGSGLTRVME---DPMIPSPFPPRPAKSILNIR 835
Qy 472 EVQLN-----GELLQWVDETIIPALHEMALAPGSTLGLPAFSYGF-----Y 512
Db 836 DTHLYVNDLIDPVTKQWKGLRQLVDPDSIDIPLI--LGIRPSRTYKSDDFSFTKSGNY 893
Qy 513 VIRN 516
Db 894 TVKS 897

RESULT 6
A48546
genome polyprotein - infectious bursal disease virus (strain A) (fragment)
C;Species: infectious bursal disease virus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-Dec-2004
C;Accession: A48546
R;Lana, D.P.; Beisel, C.E.; Silva, R.F.
Virus Genes 6, 247-259, 1992
A;Title: Genetic mechanisms of antigenic variation in infectious bursal disease virus: ar
A;Reference number: A48546; MUID:93033176; PMID:1329340
A;Accession: A48546
A;Molecule type: genomic RNA
A;Residues: 1-503 <LAN>
A;Cross-references: UNIPROT:Q04094; UNIPARC:UPI00001786F9
A;Note: sequence extracted from NCBI backbone (NCBIN:114640, NCBI:114641)
C;Genetics:
A;Map position: segment A
C;Keywords: polyprotein

Query Match 3.9%; Score 106; DB 2; Length 503;
Best Local Similarity 24.1%; Pred. No. 1.5;
Matches 69; Conservative 46; Mismatches 109; Indels 62; Gaps 16;
QY 232 LSOHPLYRHAELVGLDVGQPRKHTQHLRLSRFMSGGKAI-----DSVTWHHYV--NG 282
DB 33 LEKHTLRSETSTNLTVD-----TGSLIVFFPGFPGSIVGAHYTLQSN 78
QY 283 RSATREDFLSPEVLD---SPATAIHDLVIGVEATVPGKVKWLGETGSA--YGGGAPQLSN 337
DB 79 NYKFDQMLLTAQNLPASNYNCRVLSRSLTVRSSTLPGGVYALNGTINAVTFQGSLSLTD 138
QY 338 TYVAGFMWL-----DKLG-LAARRGIDVVMRQVSFGAGSYHLVDAGPKPLPDYWLSLLYK 391
DB 139 VSYNGLMSATANINDKIGNLVGEGVTLSLTPSGVYALNGTINAVTFQGSLSLTD 138
QY 392 RLVGTRVLQASVEQADARRPRVYLHCTNPRH---PKYREGDV--TLFALNLSNVTQSLQL 446
DB 192 KMV-----ATCSDS--RPRVYTTAADDYQFSSQYQGGVTITLFSANIDAIT--SLSV 242
QY 447 PKQLWSKSDVQYLLLP-----GKD--SILSRVQLNGRLQWVDD 485
DB 243 GSELVFKTSVQSLVIGATIYILIGDGTAVITRAVAANGLTAGIDN 288

RESULT 7

S71934
genome polyprotein - infectious bursal disease virus (strain E/DEL) (fragment)
N/Contains: major structural protein VP2; nonstructural protein VP4
C/Species: Infectious bursal disease virus
A/Variety: strain E/DEL
C/Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C/Accession: S71934; S60291
R/Vakharina, V.N.
submitted to the EMBL Data Library, October 1990
A/Reference number: S71934
A/Accession: S71934
A/Molecule type: genomic RNA
A/Residues: 1-473 <VAK>
A/Cross-references: UNIPROT:Q64957; UNIPARC:UPI00000F6949; EMBL:X54858; NID:g486622; PID
A/Experimental source: strain E/DEL
R/Vakharina, V.N.; Ahamed, B.; He, J.
Avian Dis. 36, 736-742, 1992
A/Title: Use of polymerase chain reaction for efficient cloning of dsRNA segments of inf
A/Reference number: S60291; MUID:93038354; PMID:1329714
A/Accession: S60291
A/Molecule type: genomic RNA
A/Residues: 206-350 <VAV>
A/Cross-references: UNIPARC:UPI00001786F6; EMBL:X54858
A/Experimental source: strain E/DEL
C/Comment: This virus is responsible for a severe immunodepressive disease in young chick

Query Match 3.8%; Score 105; DB 2; Length 473;
Best Local Similarity 24.1%; Pred. No. 1.6;
Matches 69; Conservative 45; Mismatches 110; Indels 62; Gaps 16;
QY 232 LSOHPLYRHAELVGLDVGQPRKHTQHLRLSRFMSGGKAI-----DSVTWHHYV--NG 282
DB 33 LEKHTLRSETSTNLTVD-----TGSLIVFFPGFPGSIVGAHYTLQSN 78
QY 283 RSATREDFLSPEVLD---SPATAIHDLVIGVEATVPGKVKWLGETGSA--YGGGAPQLSN 337
DB 79 NYKFDQMLLTAQNLPASNYNCRVLSRSLTVRSSTLPGGVYALNGTINAVTFQGSLSLTD 138
QY 338 TYVAGFMWL-----DKLG-LAARRGIDVVMRQVSFGAGSYHLVDAGPKPLPDYWLSLLYK 391

DB 139 VSYNGLMSATANINDKIGNLVGEGVTLSLTPSGVYALNGTINAVTFQGSLSLTD 138
QY 392 RLVGTRVLQASVEQADARRPRVYLHCTNPRH---PKYREGDV--TLFALNLSNVTQSLQL 446
DB 192 KMV-----ATCSDS--RPRVYTTAADDYQFSSQYQGGVTITLFSANIDAIT--SLSV 242
QY 447 PKQLWSKSDVQYLLLP-----GKD--SILSRVQLNGRLQWVDD 485
DB 243 GSELVFKTSVQSLVIGATIYILIGDGTAVITRAVAANGLTAGIDN 288

RESULT 8

GNXSIE
genome polyprotein - infectious bursal disease virus (strain E) (fragment)
N/Contains: major structural protein VP2; nonstructural protein VP4
C/Species: Infectious bursal disease virus
A/Note: host Gallus gallus (chicken)
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C/Accession: PQ0283
R/Heine, H.G.; Haritou, M.; Failla, P.; Fahey, K.; Azad, A.
J. Gen. Virol. 72, 1835-1843, 1991
A/Title: Sequence analysis and expression of the host-protective immunogen VP2 of a vari
A/Reference number: PQ0283; MUID:91341469; PMID:1651980
A/Accession: PQ0283
A/Molecule type: genomic RNA
A/Residues: 1-496 <HEI>
A/Cross-references: UNIPROT:P29802; UNIPARC:UPI00000615FC; GB:D10065; NID:g221889; PIDN:
C/Comment: This virus is responsible for a severe immunodepressive disease in young chick
C/Genetics:
A/Map position: segment A
C/Superfamily: birnavirus segment A genome polyprotein
C/Keywords: nonstructural protein; polyprotein; structural protein
F/1-452/Product: structural protein VP2 #status predicted <VP2>
F/453-496/Product: nonstructural protein VP4 (fragment) #status predicted <VP4>

Query Match 3.8%; Score 105; DB 1; Length 496;
Best Local Similarity 24.1%; Pred. No. 1.8;
Matches 69; Conservative 45; Mismatches 110; Indels 62; Gaps 16;
QY 232 LSOHPLYRHAELVGLDVGQPRKHTQHLRLSRFMSGGKAI-----DSVTWHHYV--NG 282
DB 33 LEKHTLRSETSTNLTVD-----TGSLIVFFPGFPGSIVGAHYTLQSN 78
QY 283 RSATREDFLSPEVLD---SPATAIHDLVIGVEATVPGKVKWLGETGSA--YGGGAPQLSN 337
DB 79 NYKFDQMLLTAQNLPASNYNCRVLSRSLTVRSSTLPGGVYALNGTINAVTFQGSLSLTD 138
QY 338 TYVAGFMWL-----DKLG-LAARRGIDVVMRQVSFGAGSYHLVDAGPKPLPDYWLSLLYK 391
DB 139 VSYNGLMSATANINDKIGNLVGEGVTLSLTPSGVYALNGTINAVTFQGSLSLTD 138
QY 392 RLVGTRVLQASVEQADARRPRVYLHCTNPRH---PKYREGDV--TLFALNLSNVTQSLQL 446
DB 192 KMV-----ATCSDS--RPRVYTTAADDYQFSSQYQGGVTITLFSANIDAIT--SLSV 242
QY 447 PKQLWSKSDVQYLLLP-----GKD--SILSRVQLNGRLQWVDD 485
DB 243 GSELVFKTSVQSLVIGATIYILIGDGTAVITRAVAANGLTAGIDN 288

RESULT 9

B82488
hypothetical protein VCA0200 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 31-Dec-2004
C/Accession: B82488
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301

RESULT 11

A69409
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain C/Species: Archaeoglobus fulgidus
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C/Accession: A69409
R/Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirk, G.Lodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.Nature 390, 364-370, 1997
A/Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing bacterium Pyrococcus furiosus
A/Reference number: A69250; PMID:98049343; PMID:9389475
A/Accession: A69409
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Cross-references: 1-1076 <KLE>
A/GeneBank accession: F09469
C/Function:
A/Pathway: glutamate metabolism; pyrimidine nucleotide biosynthesis
C/Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain C/Keywords: ligase; pyrimidine nucleotide biosynthesis
F/9-1055/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain F/9-469/Domain: biotin carboxylase homology <BC1>
F/562-994/Domain: biotin carboxylase homology <BC2>

	Query Match	3.7%	Score 101;	DB 2;	Length 1076;
	Best Local Similarity	20.7%;	Pred. No. 12;		
	Matches	75;	Conservative 49;	Mismatches 136;	Indels 102; Gaps 14
Qy	186	LLGCAQRSYNI	SWELGNEPNS	FRKSGICIGDGFQ	LGRDVFVHLROLLSOHPLYRHAELYG 245
		: : : :	: : : :	: : : :	: : : :
Db	640	IVQFGGQTLN	IAREL---	EDSGARILGTS	VDSIDIAEDRERFAELLER----- 685
		: : : :	: : : :	: : : :	: : : :
Qy	246	LDVGQPRKHTQ	-----	LLRSFMKSGGKAI	----DSVTWHHYVNGRSA 285
		: : : :	: : : :	: : : :	: : : :
Db	686	LNTPQPPNG	IAHSLSEAEK	IAIKTGFPVLV	APSPYVLGGRAMEIVVDTEETLERYITEALEV 745
		: : : :	: : : :	: : : :	: : : :
Ov	286	TREDFLSPEV	LDSPATAIHD	VLGI-VEATVP	GKKVWLG-----ETGSAYGGGAPQLSN 337
		: : : :	: : : :	: : : :	: : : :

Db 746 SPE---KPIIIDKF---LEDAIEVEDALCDGEVWVIGGIMEHIEEAGVHSGD\$A----- 794

QY 338 TYVAGFWMLDKGLAARGIDVVMRQVSGFAGSYHLVDAGFKPLDPDWLSL-LYKRLVGT 396

Db 795 -----CVLPPVS-----LDEVINTIVDTRKLLALNALNVGL 826

QY 397 RVULOASVEQADARRPRVYLHCTNPRHPKYREGDVTLPALNLSNVTQSLQPLPKQLWKSVD 456

Db 827 INIQYAVKDG-----KYVLEANPRASRTVPFVSKATGIPLAKIAAKLMMGKKLRELGVK 881

QY 457 QYLLLPH-----GKDSILSRVQLNGRLQLQWDDETLPALHEMALAPG\$TL 502

Db 882 EKULKHVAKEAVFPPIKLPGVDPVLGPEMKSTGEVNGIDYDFGL-AYYKAELAA\$GKL 940

QY 503 GL 504

Db 941 PL 942

RESULT 12

D97029

ribonucleotide reductase, vitamin B12-dependent [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C;Accession: D97029

; R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: D97029

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-987 <KUR>

A;Cross-references: UNIPROT:Q97K72; UNIPARC:UPI00000CA083; GB:AE001437; PIDN:AAK79023.1;

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC1047

Query Match 3.6%; Score 99; DB 2; Length 987;

Best Local Similarity 18.2%; Pred. No. 15;

Matches 65; Conservative 52; Mismatches 112; Indels 128; Gaps 14;

QY 125 LQKLLLAESHKWKHKNTTITRSTLDLHTFASSS-GFRLVFGNLALLRRAGLQWDSSNA 183

Db 442 LVQLLEASHWGQPTEDVART-----HMFRTIGLQ---ISNLAALLVMVNGPYDSEEG 494

QY 184 K-----QLLGYCAQRSYNISWELGNEPNSFRKKSIGICIDGFGQGRDFVHLRQLLSQHPLXR 239

Db 495 RAISSSLIGVLGTGYSYISSLMAKEVGAPEK----- 525

QY 240 HAELYGLDVQPKRHTQHLRLSPMKSGGK---AIDSVTHHHYVNGRSATRED--FLSPE 294

Db 526 ----YNIN----KEHMLKVLNR\$ARAAGSIDTPEKIGYKPLVNVNHSILEKEDLRVISSE 577

QY 295 VLDSFATAIHDLVIGIVEATVPGKKVLGEGTSGAYGGAPOLSNVTYVAGFWMLDKLGLAAR 354

Db 578 LKNSMNSAL-----ESGEKYGFRNAQVSNAP\$TG----- 606

QY 355 RGIDVVMRQVSGFAGSYHLVDAGFKPLDPDWLSLKYRLVGT\$RVULOASVEQADARRPRV 414

Db 607 -----TISFA-----MDCAST\$VEPF\$HVIYKLSGGG-----Y 636

QY 415 LHCTNP-----RHPKYREGDVTLPALNLSNVTQSLQPLPKQLWKSVDQYLLPHGK 465

Db 637 MTLTNPLOESLKHLYGYSENEI-----EDITNYILRKKKVKTDDCGTEYETILDGK 686

RESULT 13

D64988

yejO protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C;Accession: D64988

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi .A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: D64988

A;Status: preliminary;

A;Molecule type: DNA

A;Residues: 1-836 <BLAT>

A;Cross-references: UNIPARC:UPI0000168114; GB:AE000308; GB:U00096; NID:g1788508; PIDN:AA

A;Experimental source: strain K-12, substrain MGI655

C;Genetics:

A;Gene: yejO

Query Match 3.6%; Score 98.5; DB 2; Length 836;

Best Local Similarity 22.5%; Pred. No. 13;

Matches 72; Conservative 36; Mismatches 109; Indels 103; Gaps 17;

QY 134 HSWKKHKN-TTITRSTLDLHTFASSSGFRVLVFGNLALLRRAGLQWDSSNAKQLIGYCAQ 192

Db 523 HSWSLAENRAQITPSTTDVLNMAA\$QP---LVFDA-----ELDTVRERLGSVKG 568

QY 193 RSYNIS-WELG-NEPNSFRKKSIGI---CIDGFGQGRDFVHLRQLLSQHPLYRHAELYL 246

Db 569 VSYDTAMWSSAINTRNNVTVDAGAGFQ\$TLTGLTGLID-----SRFSREES\$TIRGL 620

QY 247 DVGQPKRHTQHLRLSPMKSGKAIDSVT-----WHHY---YVNGRSATRED\$LSPEVLD 297

Db 621 IFGY-----SHSDIGFDRGKGKGNIDSYTLGAYAGMEHQNGAYVDG-----VWKVD 665

QY 298 SFATAIHDLVIGIVEATVPGKKVWLGETSGAYGGAPOLSNVTYV-AGFWMLDKL----- 349

Db 666 RFANTH-----GK---MSNGATAFGDYN\$NGAGAHV\$SGFRWVDGL\$SVRPL 711

QY 350 -----GLAARRGIDVVMRQVSGFAGSYHLVDAGFKPLDPDWLSLKYRL 393

Db 712 AFTGFTTGDQGYTL\$N\$GMRADVGNTRILRAEAGTAV\$YHM-DLQNGTTLEPWLKAAV\$RQ- 769

QY 394 VGT\$RVULOASVEQADARRPRV 413

Db 770 -----EYADSNQV\$K 779

RESULT 14

JS0652

aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) - Streptomyces griseus (strain SS-119f

C;Species: Streptomyces griseus

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004

C;Accession: JS0652

R;Ishikawa, J.; Hotta, K.

Gene 108, 127-132, 1991

A;Title: Nucleotide sequence and transcriptional start point of the kan gene encoding an

A;Reference number: JS0652; MUID:92104494; PMID:1761222

A;Accession: JS0652

A;Molecule type: DNA

A;Residues: 1-284 <ISH>

A;Cross-references: UNIPROT:Q54216; UNIPARC:UPI00000B0D8B; GB:D00681; NID:g217013; PIDN:I

C;Comment: This enzyme confers kanamycin resistance.

C;Genetics:

A;Gene: kan

C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

C;Keywords: acyltransferase; coenzyme A

Query Match 3.6%; Score 97.5; DB 2; Length 284;

Best Local Similarity 26.8%; Pred. No. 3.2;

Matches 57; Conservative 17; Mismatches 82; Indels 57; Gaps 12;

QY 304 HDVLGIVEATVPGKKVWLGETSGAYG--GGAPOLSNVTYVAGFWMLDKLGLAARGIDVVM 361

Db 21 HDLAAL--GLVPGDVTMFHTRLSAIGYVSGPQTV-----IDAL-----LDVV- 61

```
QY 362 ROVSFGAGSYHLVDAGFKPLPDY-----WLSLLYKRLVGTIVLQASVEQADARRPV 413
Db 62 -----GPTGTLVTCGWNADAPPDYFTWPPAQEAIVRAHHAFDPRTRSEAHANGRLPEA 116
QY 414 YLH---CTNPRHPKYREGDVTLFALNLSNVTQSLQPLQKLSKSVQDVQYLLPHGKDSILS 470
Db 117 LRRRPGAVRSRHP-----DVSIALGAS-----APALMDAHPWDD-----PHGPGSPLA 160
QY 471 REVOLNORLLOM-VDDETLPALHE---MALAPG 499
Db 161 RLVALGGRVLLGAPRDTMTLLHAEALQAAPG 193

RESULT 15
S44560
alpha,alpha-trehalase (EC 3.2.1.28) NTH2 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBR001c; protein YBR0106
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jun-1994 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: S44560; S45852; S37322
R;Wolfe, K.H.; Lohan, A.J.E.
Yeast 10 (Suppl.A), S41-S46, 1994
A;Title: Sequence around the centromere of Saccharomyces cerevisiae chromosome II: similar
A;Reference number: S44556; MUID:94378721; PMID:8091860
A;Accession: S44560
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-780 <WOL>
A;Cross-references: UNIPROT:P35172; UNIPARC:UPI0000052EA3; EMBL:Z26494; NID:9403311; PID
R;Lohan, A.J.E.; Wolfe, K.H.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45730
A;Accession: S45852
A;Molecule type: DNA
A;Residues: 1-780 <LOH>
A;Cross-references: UNIPARC:UPI0000052EA3; EMBL:Z35870; NID:9536185; PIDN:CAA84937.1; PI
C;Genetics:
A;Gene: SGD:NTH2
A;Cross-references: SGD:S0000205; MIPS:YBR001c
A;Map position: 2R
C;Keywords: glycosidase; hydrolase

Query Match 3.6%; Score 97.5; DB 2; Length 780;
Best Local Similarity 22.0%; Pred. No. 14;
Matches 75; Conservative 49; Mismatches 110; Indels 107; Gaps 19;

QY 217 DGPQLGRDFVHLRQLLSQHL---YRHAELYGLDVGPQRKHTQ-HLLR-----S 261
Db 39 EGQGGRR--HRRLLSMHEYDFPFSNAEVYGPITDPKQSKIHLRLNRTMTSVFNKVS 96
QY 262 FMKSGGKAIDSVTWHYYVNGRSATREDELSPE-----VLDSFATAIHVLG----- 308
Db 97 DFKNGMK-----DYTLKRGSEDDSFSSQGNRRFPYIDNVDLALDELLASEDTDKNH 148
QY 309 --IVEATVPKGVMLGTSGAYGGAPQLSNVTYVAGFWMLDKGLAARRGIDVVMQVVSF 366
Db 149 QITIEDT--GPKVIKVGTSANGFKNVNRCTYMLSNL-LQELTIK-----SF 194
QY 367 GAGSYHLVDAGFKPLP-----DYWLSL-----LYKRLVGTIVLQASVEQADARRP 411
Db 195 GRHQIFLDEARINENPVDRLSRLITTFWTSLTRRVDLYN--IAEIARDSKIDTPGAKNP 252
QY 412 RVYL--HCTNPRHPKYREGDVTLFALNLSNVTQSLQ-----LPKQLWSKSVQDVQYLLPHG 464
Db 253 RIYVPYNC-----PEQYE-----FYIQASQMNPSLKLEVEYLPKIDITAEYVYKSLNDTP-- 300
QY 465 KDSILSREVQLNGVLLQWVDDETLPALHEMALPGSTIGLP 505
Db 301 -----GLLALAMEHVNPNSTGERSL-----VGYP 324
```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 108.256 Seconds

(without alignments)
2293.354 Million cell updates/sec

Title: US-10-645-659A-4

Perfect score: 2841

Sequence: 1 MLLRSKPALPPPLMLLLGP.....LPAFSYFFVIRNAKVAACI 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.8.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*
- 9: Geneseq2005s.*
- 10: Geneseq2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	543	2	AAY17082 Human hep
2	2841	100.0	543	4	AAB86206 Human hep
3	2841	100.0	543	7	ADD18950 Human dis
4	2841	100.0	543	8	ADK52086 Human ato
5	2841	100.0	543	8	ADM48759 Human hpa
6	2841	100.0	543	8	ADM05074 Antipsori
7	2841	100.0	543	8	ADM04902 Antipsori
8	2841	100.0	543	8	ADQ80372 Heparanas
9	2841	100.0	543	8	ADR88210 Human pre
10	2841	100.0	543	8	ADP25079 PRO poly
11	2841	100.0	543	8	ADT78177 Human hep
12	2841	100.0	543	9	ADY27036 Human hep
13	2841	100.0	543	9	AEA42426 Human hep
14	2841	100.0	588	2	AAY30124 A human p
15	2838	99.9	543	2	AAY02345 A human h
16	2838	99.9	543	3	AAY57590 Human hep
17	2838	99.9	543	3	AAB08849 Amino aci
18	2838	99.9	543	3	AAY52990 Human hep
19	2838	99.9	543	4	AAY97635 Human hep
20	2838	99.9	543	5	ABB07813 Human hep
21	2838	99.9	543	7	ADG88800 Human hpa
22	2838	99.9	543	8	ADL16379 Human hep
23	2838	99.9	543	8	ADM48716 Human hpa

24	2838	99.9	543	9	AEA42466	Aea42466 Human hep
25	2838	99.9	543	10	AE966848	Ae966848 Human hep
26	2838	99.9	592	2	AAY02346	Aay02346 A human h
27	2838	99.9	592	3	AAB08850	Aab08850 Amino aci
28	2838	99.9	592	7	ADG88804	Adg88804 Human SK-
29	2838	99.9	592	8	ADL16383	Adl16383 Human hep
30	2838	99.9	592	8	ADM48720	Adm48720 Human SK-
31	2838	99.9	592	9	AEA42461	Aea42461 Human hep
32	2835	99.8	543	8	ADO63831	Ado63831 Human hep
33	2835	99.8	543	8	ADO63823	Ado63823 Human hep
34	2835	99.8	543	8	ADO63832	Ado63832 Human hep
35	2835	99.8	543	8	ADO63822	Ado63822 Human hep
36	2829	99.6	543	4	AAB88361	Aab88361 Human mem
37	2829	99.6	543	8	ADO63824	Ado63824 Human hep
38	2829	99.6	543	9	ADY63087	Ady63087 Human clo
39	2824.5	99.4	556	9	ADZ19010	Adz19010 Heparanas
40	2820	99.3	545	6	ABP56822	Abp56822 Human hep
41	2820	99.3	545	7	ADE16012	Adel6012 G-coupled
42	2820	99.3	545	8	ADL93951	Adl93951 Human G-c
43	2807.5	98.8	570	9	ADZ19008	Adz19008 Heparanas
44	2767	97.4	530	2	AAY34173	Aay34173 Human pre
45	2740	96.4	532	2	AAY17083	Aay17083 Seq ID No

ALIGNMENTS

RESULT 1
AAY17082
ID AAY17082 standard; protein; 543 AA.
XX
AC AAY17082;
XX
DT 21-JUL-1999 (first entry)
XX
DE Human heparanase enzyme.
XX
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.
XX
OS Homo sapiens.
XX
FN WO9921975-A1.
XX
PD 06-MAY-1999.
XX
PF 28-OCT-1998; 98WO-AU000898.
XX
PR 28-OCT-1997; 97AU-00000062.
PR 09-DEC-1997; 97AU-00000812.
XX
PI (AUSU) UNIV AUSTRALIAN NAT.
XX
PR Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
XX
WPI: 1999-312956/26.
DR N-PSDB; AAX37259.
XX
PT Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing.
XX
PS Claim 6; Page 69-73; 112pp; English.
XX
CC The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a human heparanase

XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
DB 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFPGTKTDFLPDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFPGTKTDFLPDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPPDVEEKLRLWPYQEOQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
DB 121 QVNQDICKYGSIPPDVEEKLRLWPYQEOQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHHYL 300
DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHHYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGAPLLSDTFA 360
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGAPLLSDTFA 360
QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPDVLWLSLLFKKLVGTQVLM 420
DB 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPDVLWLSLLFKKLVGTQVLM 420
QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKSGDLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
DB 421 ASVQGSRRKRLRVYLHCTNTDNPRYKSGDLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKQWDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVQLNGLTLKQWDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540

RESULT 2

AAB86206
ID AAB86206 standard; protein; 543 AA.

XX AC

XX AC

XX AC

DT 24-AUG-2001 (first entry)

XX Human heparanase inhibitor protein.

DE Heparanase; inhibitor; cardiac insufficiency; cardiatic; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human.

XX Homo sapiens.

OS DE19955803-A1.

XX

PN

XX

XX

PD 23-MAY-2001.

XX 19-NOV-1999; 99DE-01055803.

XX 19-NOV-1999; 99DE-01055803.

XX (KNOL) KNOLL AG.

XX Herr D, Hahn A, Laux V;

XX WPI; 2001-368371/39.

DR N-PSDB; AAH20940.

XX

PT Treatment or prevention of cardiac insufficiency and related conditions,

PT e.g. pulmonary congestion and dyspnoea, comprises administration of

PT heparanase inhibitor.

XX

PS Disclosure; Page 11-13; 16pp; German.

XX

CC This invention describes a novel heparanase inhibitor which can be used

CC for the treatment or prevention of cardiac insufficiency and associated

CC indications, symptoms and/or malfunctions. The heparanase inhibitor of

CC the invention has cardiant, nephrotropic and hepatotropic activity. The

CC products of the invention can be used in human and veterinary medicine,

CC for the treatment or prevention of congestive heart failure e.g. primary

CC cardiomyopathy. Associated conditions treated or prevented with the

CC inhibitor are especially peripheral odemas, pulmonary and hepatic

CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.

CC nocturia can also be treated. This sequence represents the human

CC heparanase protein described in the method of the invention

XX

SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 4; Length 543;

Best Local Similarity 100.0%; Pred. No. 3.2e-275;

Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60

DB 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60

QY 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFPGTKTDFLPDPKKESTFEERSYQWS 120

DB 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFPGTKTDFLPDPKKESTFEERSYQWS 120

QY 121 QVNQDICKYGSIPPDVEEKLRLWPYQEOQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180

DB 121 QVNQDICKYGSIPPDVEEKLRLWPYQEOQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180

QY 181 GLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240

DB 181 GLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240

QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHHYL 300

DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHHYL 300

QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGAPLLSDTFA 360

DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGAPLLSDTFA 360

QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPDVLWLSLLFKKLVGTQVLM 420

DB 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPDVLWLSLLFKKLVGTQVLM 420

QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKSGDLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480

DB 421 ASVQGSRRKRLRVYLHCTNTDNPRYKSGDLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480

QY 481 RPLGPHGLLSKSVQLNGLTLKQWDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540

DB 481 RPLGPHGLLSKSVQLNGLTLKQWDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540

QY	541	ACI	543	61	IDANLATDPRFLIILGSPKLTARGLSPAYLRFGTGTDTFLIFDPKKESTTEERSYWQS	120
Db	541	ACI	543	61	IDANLATDPRFLIILGSPKLTARGLSPAYLRFGTGTDTFLIFDPKKESTTEERSYWQS	120
QY	ADD18950			121	QVNDQICKYGSIPDPVEEKLRLLEWYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS	180
Db	ADD18950 standard; protein; 543 AA.			121	QVNDQICKYGSIPDPVEEKLRLLEWYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS	180
AC	ADD18950;			181	GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS	240
XX				181	GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS	240
DT	15-JAN-2004 (first entry)			241	QLGEDFIOLHKLKRSKTFKNKLYGPDVGQPRRTAKMLKSFKAGGEVIDSVTWHHYL	300
DE	Human disease related protein SeqID439.			241	QLGEDFIOLHKLKRSKTFKNKLYGPDVGQPRRTAKMLKSFKAGGEVIDSVTWHHYL	300
XX	human; disease state; cytostatic; antiinflammatory; ophthalmological;			301	NGRTATREDFLNPVDLDFISSVQKVFQVVESTRPGKVKWLGETSSAYGGAPLLSDTFA	360
KW	antiarteriosclerotic; vulnery; gene therapy;			301	NGRTATREDFLNPVDLDFISSVQKVFQVVESTRPGKVKWLGETSSAYGGAPLLSDTFA	360
KW	hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;			361	AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYMLSLFLKVLGVKVL	420
KW	inflammation; erythropoiesis; glycolysis; gluconeogenesis;			361	AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYMLSLFLKVLGVKVL	420
KW	glucose transportation; catecholamine synthesis; iron transport;			421	ASVOGSKRRKLRVYLHCTNTDNPYKEGDLTYAINLHNVTKYLRPYDFSNKQVDKYL	480
KW	nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;			421	ASVOGSKRRKLRVYLHCTNTDNPYKEGDLTYAINLHNVTKYLRPYDFSNKQVDKYL	480
KW	retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;			481	RPLGPHGLLSKSVOLNGLTLKMWDDOTLPLMEKPLRPGSSIGLPAFSYSFFVIRNAKVA	540
KW	inflammatory condition; wound healing.			481	RPLGPHGLLSKSVOLNGLTLKMWDDOTLPLMEKPLRPGSSIGLPAFSYSFFVIRNAKVA	540
XX	Homo sapiens.			541	ACI	543
XX	WO2003018621-A2.			541	ACI	543
PN	06-MAR-2003.					
PD	23-AUG-2002; 2002WO-GB003892.					
XX	23-AUG-2001; 2001GB-00020558.					
XX	05-OCT-2001; 2001GB-00024037.					
PR	(OXFO-) OXFORD BIOMEDICA UK LTD.					
XX	Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;					
PI	WPI; 2003-290046/28.					
XX	N-PSDB; ADD18951.					
DR	New substantially purified polypeptide, useful for diagnosing or treating					
XX	a hypoxia-regulated condition, such as cancer, ischemia, reperfusion					
PT	injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or					
PT	wound healing.					
XX	Claim 25; SEQ ID NO 439; 424pp; English.					
PS	This invention relates to novel human genes and gene product which are					
XX	implicated in certain disease states. Compounds which modulate the					
CC	proteins of the invention may have cytostatic, antiinflammatory,					
CC	ophthalmological, antiarteriosclerotic or vulnery activities. The					
CC	sequences of the invention may be useful for gene therapy. The invention					
CC	may be useful for diagnosing or treating a hypoxia-regulated condition,					
CC	such as tumorigenesis, angiogenesis, apoptosis, inflammation,					
CC	erythropoiesis, or the biological response to hypoxia conditions					
CC	including processes such as glycolysis, gluconeogenesis, glucose					
CC	transportation, catecholamine synthesis, iron transport or nitric oxide					
CC	synthesis. The disease includes cancer, ischaemic conditions, reperfusion					
CC	injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,					
CC	inflammatory conditions or wound healing. The present sequence is that of					
CC	a disease related protein of the invention.					
XX	Sequence 543 AA;					
SQ	Query Match 100.0%; Score 2841; DB 7; Length 543;					
	Best Local Similarity 100.0%; Pred. No. 3.2e-275;					
	Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MLLRSKALPPPLMLLLGLPLSPGALPRPAQDVVDLDFFTQBLHLVSPSFLSVT	60			
Db	1	MLLRSKALPPPLMLLLGLPLSPGALPRPAQDVVDLDFFTQBLHLVSPSFLSVT	60			

Detecting atopic dermatitis or psoriasis comprises assaying levels of expression of an indicator gene at a rash site and non-rash site of a

PT person with atopic dermatitis or psoriasis.
XX Example 2; SEQ ID NO 119; 484pp; Japanese.
XX
CC The invention relates to detecting atopic dermatitis or psoriasis
CC comprising assaying the levels of expression of an indicator gene at a
CC rash site and non-rash site of a person with atopic dermatitis or
CC psoriasis, comparing these levels with those of a healthy person, and
CC determining that if the levels of indicators are higher or lower, then
CC this indicates the disease. Also included are a reagent for detecting
CC atopic dermatitis or psoriasis, a kit for screening for treatments, a
CC transgenic non human vertebrate animal models for the diseases, an agent
CC for inducing the diseases in mice and a DNA chip for assaying for the
CC indicator genes. The method is used for treatment, detection and animal
CC models for research of atopic dermatitis and psoriasis. The present
CC sequence is a protein encoded by an indicator gene of the invention.
XX
XX Sequence 543 AA;
SQ

Query Match 100.0%; Score 2841; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPFLSVT 60
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPFLSVT 60

Qy 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120

Qy 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNYSRSSVDVLYTFANCS 180
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNYSRSSVDVLYTFANCS 180

Qy 181 GLDLIFGNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNPNFLKKADIFINGS 240
Db 181 GLDLIFGNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNPNFLKKADIFINGS 240

Qy 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGOPRRKTAQMLKSLFKAGGEVDSVTWHYYL 300
Db 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGOPRRKTAQMLKSLFKAGGEVDSVTWHYYL 300

Qy 301 NGRTATREFLNPDLVDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360
Db 301 NGRTATREFLNPDLVDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360

Qy 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDYWLSLLFKLVGTKVLM 420
Db 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDYWLSLLFKLVGTKVLM 420

Qy 421 ASVQSGRRKRLRYLHCTNTDNPYKEGDLTLVAINLHNTKYLRLLPYPSNKKQVDKYL 480
Db 421 ASVQSGRRKRLRYLHCTNTDNPYKEGDLTLVAINLHNTKYLRLLPYPSNKKQVDKYL 480

Qy 481 RPLGPHGLLSKSVOLNGLTLKMVDDQTLPLMEKPLRPGSSLGPAPSYFFVIRNAKVA 540
Db 481 RPLGPHGLLSKSVOLNGLTLKMVDDQTLPLMEKPLRPGSSLGPAPSYFFVIRNAKVA 540

Qy 541 ACI 543
Db 541 ACI 543

RESULT 5
ID ADM48759 standard; protein; 543 AA.
XX
AC ADM48759;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human hpa protein #2.

XX Transgenic animal; heparanase; cancer; viral infection; restenosis;
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
KW human.
XX Homo sapiens.
OS
XX US2003217375-A1.
PN
XX 20-NOV-2003.
PD
XX 24-FEB-2003; 2003US-00371218.
PF
XX 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 06-FEB-2001; 2001US-00776874.
PR 19-NOV-2001; 2001US-00988113.
XX
XX (ZCHA/) ZCHARIA E.
PA (VLOD/) VLODAVSKY I.
PA (METZ/) METZGER S.
PA (PECK/) PECKER I.
PA (ILAN/) ILAN N.
PA (CHAJ/) CHAJEK-SHAUL T.
PA (GOLD/) GOLDSHMIDT O.
XX
PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
PI Chajek-Shaul T, Goldshmidt O;
XX
XX WPI; 2004-021918/02.
DR N-PSDB; ADM48748.
XX
XX New transgenic non-human animal expressing heparinase, useful as models
PT for human disease, such as cancers, viral infection, neurodegenerative
PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
XX Example 10; Fig 16; 106pp; English.
XX
CC The present invention relates to a transgenic non-human animal whose
CC genome comprises an exogenous polynucleotide sequence, including a
CC promoter active in tissues of the non-human, a region encoding a human
CC heparanase, where the promoter and the region encoding human heparanase
CC are operably linked in the exogenous polynucleotide such that human
CC heparanase is expressed in at least a portion of the cells of the non-
CC human animal. The methods and compositions of the present invention are
CC useful for the production of transgenic animals expressing heparanase, to
CC be used as models for human diseases such as cancers, viral infection,
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
CC disorders. The present sequence is human hpa protein used in the
CC exemplification of the invention.
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPFLSVT 60
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPFLSVT 60

Qy 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120

Qy 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNYSRSSVDVLYTFANCS 180
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNYSRSSVDVLYTFANCS 180

Qy 181 GLDLIFGNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNPNFLKKADIFINGS 240
Db 181 GLDLIFGNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNPNFLKKADIFINGS 240

QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTHHYL 300
| | | | |
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTHHYL 300
| | | | |
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360
| | | | |
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360
| | | | |
QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTVKVL 420
| | | | |
Db 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTVKVL 420
| | | | |
QY 421 ASVOGSKRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNVTKYLRLPYPSNKOVDKYL 480
| | | | |
Db 421 ASVOGSKRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNVTKYLRLPYPSNKOVDKYL 480
| | | | |
QY 481 RPLGPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLRPGSSGLPFAFSYFFVIRNAKVA 540
| | | | |
Db 481 RPLGPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLRPGSSGLPFAFSYFFVIRNAKVA 540
| | | | |
QY 541 ACI 543
| | |
Db 541 ACI 543
| | |
RESULT 6
ADN05074
ID ADN05074 standard; protein; 543 AA.
AC ADN05074;
XX
DT 01-JUL-2004 (first entry)
XX
DE Antipsoriatic protein sequence #716.
XX
KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS Homo sapiens.
XX
PN WO2004028479-A2.
XX
PD 08-APR-2004.
XX
PF 25-SEP-2003; 2003WO-US030907.
XX
PR 25-SEP-2002; 2002US-0414006P.
XX
PA (GETH) GENENTECH INC.
XX
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
XX WPI; 2004-305105/28.
DR N-PSDB; ADN05073.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX
XX Claim 9; SEQ ID NO 1468; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polypeptides of the invention.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 2841; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLHVSFSLSVT 60
| | | | |
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLHVSFSLSVT 60
| | | | |
QY 61 IDANLATDPRFLIILGSPKLRTLARGSLPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120
| | | | |
Db 61 IDANLATDPRFLIILGSPKLRTLARGSLPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120
| | | | |
QY 121 QVNODICKYGSIPDPVEEKRLLEWYQEQLLRLREHYQKFKFNSTYSRSSVDVLYTFANCS 180
| | | | |
Db 121 QVNODICKYGSIPDPVEEKRLLEWYQEQLLRLREHYQKFKFNSTYSRSSVDVLYTFANCS 180
| | | | |
QY 181 GLDLIFGLNALRLTADLQWNSNAQLLDYCSSKGYNISWELGNENPSFLKKADIFINGS 240
| | | | |
Db 181 GLDLIFGLNALRLTADLQWNSNAQLLDYCSSKGYNISWELGNENPSFLKKADIFINGS 240
| | | | |
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTHHYL 300
| | | | |
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTHHYL 300
| | | | |
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360
| | | | |
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360
| | | | |
QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTVKVL 420
| | | | |
Db 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTVKVL 420
| | | | |
QY 421 ASVOGSKRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNVTKYLRLPYPSNKOVDKYL 480
| | | | |
Db 421 ASVOGSKRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNVTKYLRLPYPSNKOVDKYL 480
| | | | |
QY 481 RPLGPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLRPGSSGLPFAFSYFFVIRNAKVA 540
| | | | |
Db 481 RPLGPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLRPGSSGLPFAFSYFFVIRNAKVA 540
| | | | |
QY 541 ACI 543
| | |
Db 541 ACI 543
| | |
RESULT 7
ADN04902
ID ADN04902 standard; protein; 543 AA.
XX
AC ADN04902;
XX
DT 01-JUL-2004 (first entry)
XX
DE Antipsoriatic protein sequence #631.
XX
KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
OS Homo sapiens.
XX
PN WO2004028479-A2.
XX
PD 08-APR-2004.
XX
PF 25-SEP-2003; 2003WO-US030907.
XX
PR 25-SEP-2002; 2002US-0414006P.
XX
PA (GETH) GENENTECH INC.
XX
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX
XX WPI; 2004-305105/28.
DR N-PSDB; ADN04901.
XX
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.

```

PT mammal.
XX Claim 9; SEQ ID NO 1296; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
XX treating psoriasis or a sequence having at least 80% identity to the
XX above sequences. The nucleic acid is useful for preparing a composition
XX for diagnosing or treating psoriasis in a mammal. This sequence
XX corresponds to one of the polypeptides of the invention.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 2841; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-275;
XX Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MLLRSKPAIPPPMLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60
DB 1 MLLRSKPAIPPPMLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSSDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSSDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300
DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSAYGGGAPLLSDTFA 360
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSAYGGGAPLLSDTFA 360
QY 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTQVLM 420
DB 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTQVLM 420
QY 421 ASVQSGKRRKRLRVYLHCTNTDNPYKEGDGLTVAINLHNVTKYLRLPYPSNKQVDKYL 480
DB 421 ASVQSGKRRKRLRVYLHCTNTDNPYKEGDGLTVAINLHNVTKYLRLPYPSNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLPMEKPLRPGSSGLGPAFSYSFFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLPMEKPLRPGSSGLGPAFSYSFFVIRNAKVA 540
QY 541 ACI 543
DB 541 ACI 543
XX
XX RESULT 8
XX ADQ80372
XX ID ADQ80372 standard; protein; 543 AA.
XX
XX AC ADQ80372;
XX
XX 21-OCT-2004 (first entry)
XX
XX Heparanase protein.
XX
XX cytostatic; epidermal growth factor receptor modulator; identification;
XX therapeutic response; cancer; EGFR; biomarker.
XX
XX Homo sapiens.
XX
XX WO2004063709-A2.

```

```

XX 29-JUL-2004.
XX
XX 08-JAN-2004; 2004WO-US000368.
XX
XX 08-JAN-2003; 2003US-0438735P.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Amler LC, Januario T;
XX
XX WPI; 2004-544114/52.
XX N-PSDB; ADQ80253.
XX
XX Identifying a mammal that will respond therapeutically to a method of
XX treating cancer comprises comparing the level of a biomarker in a mammal
XX before and after exposure to an epidermal growth factor receptor (EGFR)
XX modulator.
XX
XX Disclosure; SEQ ID NO 144; 520pp; English.
XX
XX The invention relates to a method of identifying a mammal that will
XX respond therapeutically to a method of treating cancer by administering
XX an epidermal growth factor receptor (EGFR) modulator by comparing the
XX level of a biomarker in a mammal before and after exposure to an EGFR
XX modulator. The method comprises: (a) measuring, in the mammal, the level
XX of at least one biomarker identified in the specification; (b) exposing
XX the mammal to the EGFR modulator; and (c) measuring in the mammal the
XX level of the biomarker, where a difference in the level in step (c)
XX compared to step (a) indicates that the mammal will respond
XX therapeutically to the method of treating cancer. The method and
XX biomarkers are useful for identifying a mammal that will respond
XX therapeutically to a method of treating cancer by administering an
XX epidermal growth factor receptor (EGFR) modulator. This sequence
XX corresponds to one of the biomarkers whose levels of expression is
XX measured in the method of the invention.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 2841; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-275;
XX Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MLLRSKPAIPPPMLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60
DB 1 MLLRSKPAIPPPMLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSSDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSSDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300
DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSAYGGGAPLLSDTFA 360
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTPGKKVWLGETSAYGGGAPLLSDTFA 360
QY 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTQVLM 420
DB 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTQVLM 420
QY 421 ASVQSGKRRKRLRVYLHCTNTDNPYKEGDGLTVAINLHNVTKYLRLPYPSNKQVDKYL 480
DB 421 ASVQSGKRRKRLRVYLHCTNTDNPYKEGDGLTVAINLHNVTKYLRLPYPSNKQVDKYL 480

```

```

Db 421 ASVQSGRRKRLVYLHCTNTDNPRYKEGDLTLVAINLHNTKYLRLPYSPFNQKQVKYLL 480
QY 481 RPLGPHGLLSKSVQLNGLTLMQVDDQTLPLMEKPLRPGSSGLGLPAPSYFFVIRNAKVA 540
    |||||
Db 481 RPLGPHGLLSKSVQLNGLTLMQVDDQTLPLMEKPLRPGSSGLGLPAPSYFFVIRNAKVA 540
QY 541 ACI 543
    |||
Db 541 ACI 543

RESULT 9
ADR88210
ID ADR88210 standard; protein; 543 AA.
AC ADR88210;
XX
XX 18-NOV-2004 (first entry)
XX Human preproheparanase.
XX
XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
KW human; heparanase; enzyme.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1. .35 /label= Signal_peptide
FT Protein 36. .543 /label= Mature_heparanase
FT Region 36. .109 /note= "8 KDa subunit of mature heparanase dimer"
FT Domain 89. .107 /note= "Functional peptide epitope"
FT Region 158. .543 /note= "45 KDa subunit of mature heparanase dimer"
FT Domain 219. .233 /note= "Functional peptide epitope"
FT Active-site 225 /note= "Active site residue"
FT Binding-site 258. .266 /note= "Putative heparin binding domain"
FT Domain 294. .307 /note= "Functional peptide epitope"
FT Domain 334. .348 /note= "Functional peptide epitope"
FT Active-site 343 /note= "Active site residue"
FT Binding-site 414. .420 /note= "Putative heparin binding domain"
FT Domain 437. .446 /note= "Functional peptide epitope"
XX
XX US2004170631-A1.
XX
XX 02-SEP-2004.
XX
XX 28-NOV-2003; 2003US-00722502.
XX
XX 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
PR 22-AUG-2003; 2003US-00645659.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.

```

```

PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VGEL/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I,
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI; 2004-625084/60.
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
XX useful for treating heparanase-associated conditions such as inflammation
XX or cancer, comprises administering a drug and an anti-heparanase antibody
XX complex.
XX
XX Claim 2; SEQ ID NO 4; 58pp; English.
XX
XX The invention relates to a method of targeted drug delivery to a tissue
XX of a patient, the tissue expressing heparanase. The method comprises
XX providing a complex of a drug directly or indirectly linked to an anti-
XX heparanase antibody, and administering the complex to the patient. In the
XX targeted drug delivery, the antibody comprises an antibody or its portion
XX capable of specifically binding to at least one epitope of a heparanase
XX protein. The composition and methods of the invention are useful for
XX diagnosing, preventing or treating conditions associated with heparanase
XX catalytic activity (e.g. an inflammatory disorder, wound, scar,
XX vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
XX proliferation, invasion of circulating tumour cells and metastatic
XX disease), for purifying heparanase, or for developing drugs for those
XX heparanase-associated conditions. The vasculopathy is atherosclerosis,
XX restenosis or aneurysm. The cancerous condition is a solid cancer or a
XX non-solid cancer. The non-solid cancer is a haematopoietic malignancy
XX selected from acute lymphocytic leukaemia (ALL), acute myelogenous
XX leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
XX leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
XX Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
XX multiple myeloma. The solid cancer is selected from tumours in lip and
XX oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
XX thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
XX anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
XX Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
XX sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
XX vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
XX tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
XX bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
XX malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
XX retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
XX brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
XX sarcoma. The present sequence is human preproheparanase.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 2841; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-275;
XX Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLVSPFLSVT 60
    |||||
Db 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLVSPFLSVT 60
    |||||
QY 61 IDANLATDPRFLILGLSPKRLTARGLSPAYLRFGGTKTDFLIDPKKESTPEERSYMQS 120
    |||||
Db 61 IDANLATDPRFLILGLSPKRLTARGLSPAYLRFGGTKTDFLIDPKKESTPEERSYMQS 120
    |||||
QY 121 QVNQDICKYGSIPDPVEEKRLWEPYQEQLLRHHYQKFKNKNSTYSRSSVDVLYTFANCS 180
    |||||
Db 121 QVNQDICKYGSIPDPVEEKRLWEPYQEQLLRHHYQKFKNKNSTYSRSSVDVLYTFANCS 180
    |||||
QY 181 GLDLIFGLNALLRTADLQWNSSNAQLLDYCSSKGYNISWELGNPNPSFLKKADIFINGS 240
    |||||

```

Db 181 GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGESSAYGGAPLLSDTFA 360
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGESSAYGGAPLLSDTFA 360
QY 361 AGFWMLDKGLSARMGIEVVNRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420
Db 361 AGFWMLDKGLSARMGIEVVNRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420
QY 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLVAINLHNVTKYLRPLYPFSNKQVDKYL 480
Db 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLVAINLHNVTKYLRPLYPFSNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
Db 481 RPLGPHGLLSKSVQLNGLTLKWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
QY 541 ACI 543
Db 541 ACI 543
RESULT 10
ADP25079 standard; protein; 543 AA.
AC ADP25079;
XX 18-NOV-2004 (first entry)
XX PRO polypeptide SEQ ID NO:2257.
KW PRO; antiinflammatory; antiarthritic; antirheumatic; immunosuppressive;
KW osteopathic; antidiabetic; dermatological; antipsoriatic; antiallergic;
KW antiasthmatic; hepatotropic; respiratory; gene therapy; immune system.
OS Unidentified.
XX WO2004041170-A2.
XX 21-MAY-2004.
XX 30-OCT-2003; 2003WO-US034312.
XX 01-NOV-2002; 2002US-0423394P.
XX (GETH) GENENTECH INC.
XX Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WT;
XX Wu TD;
XX WPI; 2004-419628/39.
XX N-PSDB; ADP25078.
XX New PRO polypeptides and polynucleotides, useful for treating e.g.
XX erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated
XX renal disease, or demyelinating diseases of the central or peripheral
XX nervous system.
XX Claim 7; SEQ ID NO 2257; 2940pp; English.
XX The invention relates to a novel isolated nucleic acid and the PRO
XX polypeptide encoded by it. A protein of the invention has
XX antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,
XX osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,
XX antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide
XX of the invention may have a use in gene therapy. The PRO polypeptide, its

CC agonist, antagonist, or antibody that specifically binds to the
CC polypeptide is useful for treating an immune related disorder such as
CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,
CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an
CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic
CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune
CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal
CC disease, a demyelinating disease of the central or peripheral nervous
CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,
CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary
CC disease, infectious or autoimmune chronic active hepatitis, primary
CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,
CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's
CC disease, an autoimmune or immune-mediated skin disease, a bullous skin
CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic
CC disease, asthma, allergic rhinitis, atopic dermatitis, food
CC hypersensitivity, urticaria, an immunologic disease of the lung,
CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity
CC pneumonitis, a transplantation associated disease, graft rejection or
CC graft-versus-host disease. The present sequence represents a PRO protein
CC of the invention.
XX
XX Sequence 543 AA;
Query Match 100.0%; Score 2841; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAOVDVLDFTORPLHLVSPFSVLT 60
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAOVDVLDFTORPLHLVSPFSVLT 60
QY 61 IDANLATDPRFLILGSPKRLTLARGSPAYLRFGTKTDFLIIDPKKESTFEERSYQWS 120
Db 61 IDANLATDPRFLILGSPKRLTLARGSPAYLRFGTKTDFLIIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSRSVDVLTFFANCS 180
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSRSVDVLTFFANCS 180
QY 181 GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
Db 181 GLDLIFGLNALRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGESSAYGGAPLLSDTFA 360
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGESSAYGGAPLLSDTFA 360
QY 361 AGFWMLDKGLSARMGIEVVNRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420
Db 361 AGFWMLDKGLSARMGIEVVNRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420
QY 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLVAINLHNVTKYLRPLYPFSNKQVDKYL 480
Db 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLVAINLHNVTKYLRPLYPFSNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
Db 481 RPLGPHGLLSKSVQLNGLTLKWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
QY 541 ACI 543
Db 541 ACI 543
RESULT 11
ADT78177
ID ADT78177 standard; protein; 543 AA.
XX

AC ADT78177;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE Human heparanase protein.
 XX
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 157..162
 FT /note= "Putative heparin binding site"
 FT Binding-site 271..277
 FT /note= "Putative heparin binding site"
 FT Binding-site 426..433
 FT /note= "Putative heparin binding site"
 FT
 XX US2004213789-A1.
 PN
 XX 28-OCT-2004.
 PD
 XX 22-AUG-2003; 2003US-00645659.
 XX
 PR 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX
 XX (YACO/) YACOBY-ZEEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (GELD/) GELDER J M V.
 PA (VL0D/) VL0DAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
 PI Friedmann Y;
 XX
 XX WPI; 2004-774790/76.
 DR
 XX
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX
 XX Claim 5; SEQ ID NO 4; 68pp; English.
 PS
 XX
 XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,

CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents human heparanase.
 XX
 SQ Sequence 543 AA;
 Query Match 100.0%; Score 2841; DB 8; Length 543;
 Best Local Similarity 100.0%; Pred. No. 3.2e-275;
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLRSKPALPPPLMLLLGPIGLSPGALPPRAQAQDVVDLDFQEPHLHVPSPFLSVT 60
 DB 1 MLLRSKPALPPPLMLLLGPIGLSPGALPPRAQAQDVVDLDFQEPHLHVPSPFLSVT 60
 QY 61 IDANLATDPRFLILGLSPKLRTLARGLSPAYLRFQGTGTDFLIFDPKKESTFEERSYWOS 120
 DB 61 IDANLATDPRFLILGLSPKLRTLARGLSPAYLRFQGTGTDFLIFDPKKESTFEERSYWOS 120
 QY 121 QVNQDICKYISIPDPVEEKLRLWPYQQLLREHYQKKFNSTYSRSSVDVLYTFANCS 180
 DB 121 QVNQDICKYISIPDPVEEKLRLWPYQQLLREHYQKKFNSTYSRSSVDVLYTFANCS 180
 QY 181 GLDLIFGLNALRLTADLQWSSNAQLLDYCSSKGYNISWELGNPNPNSFLKKADIFINGS 240
 DB 181 GLDLIFGLNALRLTADLQWSSNAQLLDYCSSKGYNISWELGNPNPNSFLKKADIFINGS 240
 QY 241 QLGEDFTQLHKLRLKSTPFNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTHHYYL 300
 DB 241 QLGEDFTQLHKLRLKSTPFNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTHHYYL 300
 QY 301 NGRTATREDPLNPDVLDIFISSVOKVFOVVESTRPCKVWLGETSSAYGGGAPLLSDTFA 360
 DB 301 NGRTATREDPLNPDVLDIFISSVOKVFOVVESTRPCKVWLGETSSAYGGGAPLLSDTFA 360
 QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKKLVGTVKVL 420
 DB 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKKLVGTVKVL 420
 QY 421 ASVQSGKRRKRLRVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKLYLPPFPFNKQVDKYL 480
 DB 421 ASVQSGKRRKRLRVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKLYLPPFPFNKQVDKYL 480
 QY 481 RPLGPHGLLSKSVQLNGTLTKQVDDQTLPLMEKPLRPGSSILGPAFYSFFVIRNAKVA 540
 DB 481 RPLGPHGLLSKSVQLNGTLTKQVDDQTLPLMEKPLRPGSSILGPAFYSFFVIRNAKVA 540
 QY 541 ACI 543
 DB 541 ACI 543
 RESULT 12
 ADY27036
 ID ADY27036 standard; protein; 543 AA.
 XX
 AC ADY27036;
 XX
 DT 05-MAY-2005 (first entry)
 XX
 DE Human heparanase protein.
 XX
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.
 XX
 OS Homo sapiens.
 XX
 PN WO2005016227-A2.

XX PD 24-FEB-2005.
XX PF 12-AUG-2004; 2004WO-11000744.
XX PR 14-AUG-2003; 2003US-0494800P.
XX PR 12-JAN-2004; 2004US-0535492P.
XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PI Van-Gelder JM, Miron D;
XX DR WPI; 2005-182203/19.
XX
XX Regulating heparanase activity, useful for treating heparanase-associated
XX PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
XX PT neurological diseases or viral diseases) comprises modulating heparanase
XX PT activation.
XX PS
XX PS Disclosure; SEQ ID NO 8; 211pp; English.
XX
XX The invention relates to a method of regulating heparanase activity in a
XX CC tissue or regulating a biological process depending at least in part on
XX CC heparanase activity comprising modulating heparanase activation. The
XX CC invention also relates to methods of treating a heparanase- or heparin
XX CC binding protein-associated disease or disorder in a subject, a
XX CC pharmaceutical composition for use in the treatment of a heparanase-
XX CC associated disease or disorder comprising a therapeutic amount of an
XX CC agent capable of modulating heparanase activation and a pharmaceutical
XX CC carrier or diluent, a method of identifying a protease activator of
XX CC heparanase, a protease substrate mimetic comprising a peptide
XX CC representing a subset or all substrate residues or cleavage sites of
XX CC human heparanase or an equivalent non-human heparanase, a method of
XX CC producing active heparanase and a method of modulating an adhesion
XX CC activity of heparanase. The composition and methods are useful for
XX CC modulating heparanase activation and for treating heparanase-associated
XX CC diseases or disorders such as cancer, inflammation, cardiovascular
XX CC diseases, neurological diseases or viral infections. This sequence
XX CC represents a human heparanase protein used in the scope of the invention.
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 2841; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLRSKPALPPMLLLGLPLGSPALPRPAQAQDVLDLDFPTQPLHLVSPFLSVT 60
DB 1 MLLRSKPALPPMLLLGLPLGSPALPRPAQAQDVLDLDFPTQPLHLVSPFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGKTDFLIDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGKTDFLIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDVEEKLRLWPYQEQLLREHYQKFKFNSTYSRSSVDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDVEEKLRLWPYQEQLLREHYQKFKFNSTYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNENPNSFLKADIFINGS 240
QY 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKMLKSLFKAGGEVDSVTHHYYL 300
DB 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKMLKSLFKAGGEVDSVTHHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360
QY 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDVLWLSLLPKLVGTVKLM 420
DB 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDVLWLSLLPKLVGTVKLM 420

QY 421 ASVQSGSRKRLRVYLHCTNTDNPYKGGDLTLYAINLHNVTYLRLPYFPFSNKQVDKYL 480
DB 421 ASVQSGSRKRLRVYLHCTNTDNPYKGGDLTLYAINLHNVTYLRLPYFPFSNKQVDKYL 480
QY 481 RPLGPHGLLSKSVOLNGITLKMVDDQTLPLMEKPLRPGSSIGLPAFSYSPFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVOLNGITLKMVDDQTLPLMEKPLRPGSSIGLPAFSYSPFVIRNAKVA 540
QY 541 ACI 543
DB 541 ACI 543
RESULT 13
AEA42426
ID AEA42426 standard; protein; 543 AA.
XX
AC AEA42426;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human heparanase protein SEQ ID NO:4.
XX
KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
KW angiogenesis disorder; cancer; tumor; metastasis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 89..107
FT Peptide /note= "heparanase epitope SEQ ID NO:7"
FT Peptide 219..233
FT Peptide /note= "heparanase epitope SEQ ID NO:8"
FT Misc-difference 246
FT Peptide /note= "encoded by TAT"
FT Peptide 294..307
FT Peptide /note= "heparanase epitope SEQ ID NO:10"
FT Peptide 334..348
FT Peptide /note
FT Peptide 437..446
FT Peptide /note= "heparanase epitope SEQ ID NO:6"
XX
XX AUZ004201462-A1.
XX
PD 06-MAY-2004.
XX
PF 08-APR-2004; 2004AU-00201462.
XX
PR 08-APR-2004; 2004AU-00201462.
XX
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinstein E;
XX
DR WPI; 2005-173343/19.
DR N-PSDB; AEA42434, AEA42435, AEA42460.
XX
XX Novel isolated antibody capable of specifically binding to epitope of
XX PT heparanase protein, useful for preventing and treating heparanase-related
XX PT disorder such as inflammatory disorder, scars, autoimmune conditions or
XX PT angiogenesis.
XX
PS Claim 2; SEQ ID NO 4; 260pp; English.
XX
XX The invention relates to an isolated antibody or its portion (I) capable
XX CC of specifically binding to an epitope of a heparanase protein. Also
XX CC described: (i) a cell line (ii) for producing a monoclonal antibody or

CC its portion, comprising a cell line for producing (1); (2) a
CC pharmaceutical composition comprising (1) and a carrier; and (3) an
CC affinity medium (III) for binding human heparanase polypeptides,
CC comprising (1) immobilized to a chemically inert, insoluble carrier. (1)
CC useful for treating a subject suffering from a pathological condition,
CC which involves administering (1) to the subject. (1) is useful for
CC preventing and treating heparanase-related disorder or condition chosen
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
CC cell proliferation, invasion of circulating tumor cells and metastatic
CC disease. (1) is useful for detecting the presence of heparanase
CC polypeptide in a sample. (1) is useful for detecting heparanase-related
CC disease or condition in a subject such as vertebrate, preferably mammal
CC e.g., human. The heparanase-related disorder or condition further
CC includes renal disease or disorder chosen from diabetic nephropathy,
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
CC and renal cell carcinoma. The present sequence represents human
CC heparanase, which is used in the exemplification of the present
CC invention.
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.2e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKALPPPLMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60
DB 1 MLLRSKALPPPLMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLPDPKKESTFEERSYWS 120
DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLPDPKKESTFEERSYWS 120
QY 121 QVNQDICKYGSIPDVEEKLREWPYQEQLLLRHYQKFKNYSRSSVDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDVEEKLREWPYQEQLLLRHYQKFKNYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFTQLHLKLRKSTFKNAKYGPDVGQPRRKTAQMLKSFLLKAGGEVDSVTWHYYL 300
DB 241 QLGEDFTQLHLKLRKSTFKNAKYGPDVGQPRRKTAQMLKSFLLKAGGEVDSVTWHYYL 300
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGETSAYGGAPLLSDTEA 360
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGETSAYGGAPLLSDTEA 360
QY 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVDFNPDLPDYWLSLLFKKLVTGTVLM 420
DB 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVDFNPDLPDYWLSLLFKKLVTGTVLM 420
QY 421 ASYQGSRRKRLRYLHCTNTDNPRYKEDGLTYLAINLHNVTYKLRPLYPFSNKQVDKYL 480
DB 421 ASYQGSRRKRLRYLHCTNTDNPRYKEDGLTYLAINLHNVTYKLRPLYPFSNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLRPGSSGLPAPSFYSFFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLRPGSSGLPAPSFYSFFVIRNAKVA 540
QY 541 ACI 543
DB 541 ACI 543

RESULT 14
AAY30124
ID AAY30124 standard; protein; 588 AA.
XX
AC AAY30124;
XX

DT 20-MAR-2003 (revised)
DT 14-OCT-1999 (first entry)
XX
DE A human protein with heparanase activity.
XX
KW Human; heparanase; heparan sulfate; trauma; autoimmune disease;
KW skin disease; cardiovascular disease; nervous system disease;
KW Alzheimer's disease; cancer; cancer metastasis; angiogenesis;
KW inflammation; arthritis.
XX
OS Homo sapiens.
XX
PN WO9940207-A1.
XX
PD 12-AUG-1999.
XX
PF 05-FEB-1999; 99WO-EP000777.
XX
PR 09-FEB-1998; 98GB-00002725.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Nakajima M, Toyoshima M;
XX
XX WPI; 1999-494300/41.
DR N-PSDB; AAX86671.
XX
PT New heparanase polypeptide useful for treating autoimmune diseases, skin
PT diseases, cardiovascular diseases and nervous system diseases including
PT Alzheimer's disease.
XX
PS Claim 3; Page 29-31; 40pp; English.
XX
CC The present sequence represents a polypeptide with human heparanase
CC biological activity. Antagonists and inhibitors of the protein prevent it
CC from degrading the extracellular matrix and releasing heparan sulfate
CC from the extracellular matrix surface. The heparanase protein or the anti
CC -heparanase antibody are used in pharmaceutical compositions for treating
CC warm blooded animals suffering from a disease resulting from shortage or
CC lack of the heparanase protein, or from excessive activity or over-
CC expression of the heparanase protein, respectively. The heparanase
CC protein is used in treating diseases such as trauma, autoimmune disease,
CC skin diseases, cardiovascular diseases and nervous system diseases
CC including Alzheimer's disease resulting from shortage or lack of
CC polypeptide. The anti-heparanase antibody is used in treating the
CC diseases like cancer, cancer metastasis, angiogenesis and inflammation
CC including arthritis, resulting from excessive activity or over expression
CC of heparanase protein. The anti-heparanase antibody can be used to detect
CC the presence or absence of polypeptide and its concentration. (Updated on
CC 20-MAR-2003 to correct PA field.)
XX
SQ Sequence 588 AA;

Query Match 100.0%; Score 2841; DB 2; Length 588;
Best Local Similarity 100.0%; Pred. No. 3.7e-275;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKALPPPLMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60
DB 46 MLLRSKALPPPLMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 105
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLPDPKKESTFEERSYWS 120
DB 106 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLPDPKKESTFEERSYWS 165
QY 121 QVNQDICKYGSIPDVEEKLREWPYQEQLLLRHYQKFKNYSRSSVDVLYTFANCS 180
DB 166 QVNQDICKYGSIPDVEEKLREWPYQEQLLLRHYQKFKNYSRSSVDVLYTFANCS 225
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
DB 226 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 285

Qy		241	QLGDEFTQLHKLLRKSTFKNAKLYGPDVQGPPRRKTAKMLKSPFLKAGGEVIDSVTWHYYL	300
Dd		286	QLGDEFTQLHKLLRKSTFKNAKLYGPDVQGPPRRKTAKMLKSPFLKAGGEVIDSVTWHYYL	345
Qy		301	NGRATATREDFLNPDVLDTIFISSVQKVQFVBESTRPGKKVWLGETSSAYCGGAPLLSDTEFA	360
Dd		346	NGRATATREDFLNPDVLDTIFISSVQKVQFVBESTRPGKKVWLGETSSAYCGGAPLLSDTEFA	405
Qy		361	AGFMWLDKLGLSARMGIEVMWRQVFPGAGNVHLVDENFDPLDPDYLSLLFKKLVGTKVJLM	420
Dd		406	AGFMWLDKLGLSARMGIEVMWRQVFPGAGNVHLVDENFDPLDPDYLSLLFKKLVGTKVJLM	465
Qy		421	ASVOGSKRRKRLRVYLHCTNTNDNPRYKEGDLTLYAINLHNVTKYLRLPYPFSNKQVDKYLL	480
Dd		466	ASVOGSKRRKRLRVYLHCTNTNDNPRYKEGDLTLYAINLHNVTKYLRLPYPFSNKQVDKYLL	525
Qy		481	RPLGPHGLLSKSQVLNGITLKXVDDQOTLPLPMEKPLRPGSSIGLPAFYSFFVIIRNAKVA	540
Dd		526	RPLGPHGLLSKSQVLNGITLKXVDDQOTLPLPMEKPLRPGSSIGLPAFYSFFVIIRNAKVA	585
Qy		541	ACI 543	
Dd		586	ACI 588	

RESULT 15

AY02345
ID AY02345 standard; protein: 543 AA.

AA
AC
AA02345;

DT	09-JUL-1999	(first entry)
XX		

DE A human heparanase protein.

Heparanase; hp; modulator; heparin-binding growth factor;
cellular response; cytokine; cell interaction; plasma lipoprotein;
cellular susceptibility; infection; disintegration;
neurodegenerative plaque; wound healing; angiogenesis; restenosis;
atherosclerosis; inflammation; neurodegenerative disease; neutralise;
plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX Homo sapiens

AA PN WO9911798-A1.

11-MAR-1999

XX
PF 31-AUG-1998: 98WO-US017954.XX
PR 02-SEP-1997: 97US-00922170.

PR 02-JUL-1998; 5803-00109388.
XX

PA (INST-) INSIGHT STRATEGI & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.

Pecker I. Vlodavsky I. Feinstein E:

WPI: 1999-302255/25.

UR N-PSUB; AAA35648.
XX

PT and inflammation.

PS Claim 6: Fig 1: 63pp; English.

CC The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of heparin-
CC binding growth factors, cellular responses to heparin-binding growth
CC factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections or

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 18.0785 Seconds
(without alignments)
2889.939 Million cell updates/sec

Title: US-10-645-659A-4
Perfect score: 2841
Sequence: 1 MLRSKPALPPPLMLLLGP.....LPASYSFFVIRNAKVACI 543
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893.5	31.5	480	JC7506	heparanase protein
2	417	14.7	521	T45608	hypothetical prote
3	169.5	6.0	190	T01953	hypothetical prote
4	112	3.9	2298	T49648	hypothetical prote
5	111	3.9	670	T10666	hypothetical prote
6	111	3.9	688	S32961	hypothetical prote
7	110.5	3.9	356	F64383	hypothetical prote
8	107.5	3.8	575	T12094	beta-fructofuranos
9	106	3.7	670	T38446	microtubule-associ
10	105.5	3.7	411	T574760	hypothetical prote
11	105.5	3.7	879	E91031	probable outer mem
12	104.5	3.7	788	S00652	phosphoribosylamin
13	104	3.7	432	F70411	adenylosuccinate s
14	104	3.7	796	D97065	transketolase [imp
15	104	3.7	2013	A11489	probable peptidogl
16	103.5	3.6	500	D87541	beta-xylosidase [i
17	103.5	3.6	676	AF1153	transcription anti
18	102	3.6	879	F85875	probable fibrinoly
19	102	3.6	897	G05259	dynein heavy chain
20	102	3.6	4644	A38905	dynein heavy chain
21	101.5	3.6	746	T46821	siderophore recept
22	101.5	3.6	746	T46821	RhtA Rhtobactin r
23	101	3.6	594	A85420	hypothetical prote
24	100.5	3.5	604	E75119	hypothetical prote
25	100.5	3.5	687	F85188	retrotransposon li
26	100.5	3.5	847	AG1001	nitrite reductase
27	100.5	3.5	1314	T31948	probable membrane
28	100.5	3.5	1734	T44101	phorbol ester-bind
29	100	3.5	578	B89045	protein B0238.7 [i

30	100	3.5	654	2	T14202	NADH2 dehydrogenas
31	99.5	3.5	587	2	S36231	beta-fructofuranos
32	99.5	3.5	989	2	A22140	toxin secretion AB
33	98.5	3.5	629	2	C64180	hypothetical prote
34	98.5	3.5	804	2	G71546	probable DNA gyras
35	98	3.4	465	2	T19113	hypothetical prote
36	98	3.4	644	2	A97268	methionyl-tRNA syn
37	98	3.4	716	1	C60008	RNA-directed RNA p
38	97.5	3.4	511	2	S61166	probable membrane
39	97	3.4	379	2	A69974	cystathionine gamm
40	97	3.4	437	1	A48061	translation releas
41	97	3.4	726	2	C86085	catalase, hydrop
42	97	3.4	726	2	G91237	hydroperoxidase HP
43	97	3.4	760	2	T34414	hypothetical prote
44	96.5	3.4	510	2	H69893	conserved hypothet
45	96.5	3.4	621	2	A95250	choline binding pr

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirtcock, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and therapi

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 31.5%; Score 893.5; DB 2; Length 480;
Best Local Similarity 35.8%; Pred. No. 8.2e-59;
Matches 201; Conservative 75; Mismatches 146; Indels 139; Gaps 9;

Qy	20	PLGLPSGAL-----PRPA-----QAQDVVDLDFPTQPLHLVSPS	55
Db	18	PPACLAFCALYIALLLHLSLSSQAGRRPLVDRAAGLUKEKTLILLDVSVKNPVRTVNE	77
Qy	56	FLSVTIDANLATDPRFLILGLSPKLTLAGLSPAYLRFGGTKTDFLIF----DPKKEST	111
Db	78	FLSLQLDPSIIHD-GWLDLFLSSKRLVTLARGUSPAFLRFGGKRTDFLOFQNLNPAKSR-	135
Qy	112	FEERSYQSQVNQDICKYISPPDVEEKRLLEWPYQEQLLLREHYQKFKPNSTSRSSVD	171
Db	136	-----GGFGPD-----YYLKNYE-----	148
Qy	172	VLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYNISWELGNEPNSFLK	231
Db	149	-----DEPNRYT	156
Qy	232	KADIFINGSQIGEDFIQLHKLRLK-STFNKAKLYGPDVGQPRKTKAKMLKSLKAGGEVI	290
Db	157	MHGRAVNGSQLGKDYIQLKSLLOPIRIYSRASLYGNIGRPRKNVIALLDGFMKVGASTV	216
Qy	291	DSVTWHYLLNGRATATREDFLNPDVLDIFISSVQKVFQVVESTRCCKVWLGETTSAYGG	350
Db	217	DAVTWQHICYIDGRVVKVWMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWLSGVVTTSAG	276
Qy	351	GAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLISLLF	410
Db	277	GTNNLSDSYAAGFLWNLTLGMLANQIGIDVVIHRSFFDHGYNHLVDQNFNPLPDYWLISLLY	336

Qy	208	LDYCSSKGYNISWELGN-----EPNSFLKKA-DIFIN-----GSQLGEDFIQLHKLL 253
Dd	197	: : : : : : : : : : : : : : : :
Db	197	--TSEGLNDSWELWNLSFRLLCEHDSKLSVALDVLSTLPSETSLGRWNMGES-VRAAIIIS 253
Qy	254	RKSITFKNAKLVPDVGOB--RRKTAKMLKSFL-KAGGEVIDSVTHWHYYLNGRTATREDF 310
Db	254	TDAFLTNAR-----GYPLCKSRHQKLTJAGFPDHAQVVVICGPVHNLOKPLDSSSEGTE 307
Qy	311	LNPVDLDTIFFISSVKVFQVWESTRPKKKVMLGCTSSAYGGGAPLLSDTFEAAFGFMWLDKLG 370
Db	308	KNP--LRILDYAVLFQKMESLSEQERIELGYRDFLOPQLPQPLMDNLSEAQTETFE--- 362
Qy	371	LSARMGIEVNMQRVFFGAGNYHLVDENFDPLPDYWLSSLFKKLGVTKVLM- 420
Db	363	--RDSVKYIQYQ---RAVEKALVDR----VPDEKASEL-----TTVLMVVGAGRGPLV 406
Qy	421	-ASVOGSKR--RKLRVYLHCNTNDNPRYKEGDILTLYAINLNHVTK-----YLRLPY 468
Db	407	RASIQAABAEEDRKLVKY---AVEKNPN-----AVVTLHNLVRMGEWDVVVTIISCDM 455
Qy	469	PFSN--KOVDKYLLRPILPGHLLSKSVQLNGTLTKMWDDOTLPLPM---EKPLRPGSSILG 523
Db	456	RFWNAPQADILVSELLSGSFG-----DNELSPCEDGAQRFLKP-DGIS 498
Qy	524	LPAFSYSFV 533
Db	499	IPS-SYTSFI 507
RESULT 6		
S32961		hypothetical protein YBR259w - yeast (<i>Saccharomyces cerevisiae</i>)
N;		Alternate names: hypothetical protein YBR1727
C;		Species: <i>Saccharomyces cerevisiae</i>
C;		Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;		Accession: S32961; S46140
R;		Doignon, F.; Biteau, N.; Crouzet, M.; Aigle, M.
Yeast 9,		189-199, 1993
A;		Title: The complete sequence of a 19,482 bp segment located on the right arm
A;		Reference number: S29348; MUID:93220397; PMID:8465606
A;		Accession: S32961
A;		Status: translation not shown
A;		Molecule type: DNA
A;		Residues: 1-688 <DOI>
A;		Cross-references: UNIPROT:P38338; UNIPARC:UPI000013A298; EMBL:X70529; NID:g1
R;		Aigle, M.; Bactet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to		The Protein Sequence Database, August 1994
A;		Reference number: S45940
A;		Accession: S46140
A;		Molecule type: DNA
A;		Residues: 1-688 <AIG>
A;		Cross-references: UNIPARC:UPI000013A298; EMBL:X70529; PIDN:CAA8A
C;		Genetics:
A;		Cross-references: SGD:S00000463
A;		Map position: 2R
C;		Superfamily: Saccharomyces cerevisiae hypothetical protein YBR259w
Query Match		3.9%; Score 111; DB 2; Length 688;
Best Local Similarity		22.5%; Pred. No. 3.6;
Matches		67; Conservative 45; Mismatches 94; Indels 92; Gaps 16
Qy	126	ICKYGSIPTDVEEKRLRWPEYQEOLLREHYQKFKNSTYGRSS-----VDVLTY 175
Db	164	MAEYSWKWSDDDRQLQFMYPFPMKLECLVKFVENFDLKQSSDPKLKELIIPWEKIVTV 223
Qy	176	PANCSGLDLIIFGLNALRTADLQNSSN-----AQILLD-----YCSSKGX 216
Db	224	-ANC--IDAFTGEQVRIDGAELIWTSKNLVFESSISSAVLRINDLNQMFSAPRPYGEALV 280
Qy	217	-----NISWELGNPNPSFLKKA--DIF--INGSOLG--EDFTQLHKLLRK----- 255
Db	281	QDFAHIRELKWDSNDKVESLIRALIFNDMPFYFNKEQVDTKADGIFFLURLRKNFKHEIN 340

Qy 107 KKESTFEERSYQWQVQNDICKYGSIPDPVEEKLRLWEPYOEQLLLREHYQKFKFNSTYS 166
Db 73 KRPSVVKSR-----KKGSENI SFMEKTKAIKQKSRREPSKFRSLARPLCITPIDSTSTPT 128
Qy 167 RSSVDVLYTFANCGLD-LIFGLNALLRTADLQWNSNAOLL-LDYCSKSGYNI SWELGN 224
Db 129 KTA---TFYTSSTTENDELNFSTEELSSFDITLLNSDTSKLSGLDDSSFMEEFEVQVDN 186
Qy 225 -----EPNSFLKK-----ADTPINGSQIGEDFIQLHKLRLK-----ST 257
Db 187 VLQCECKKFTPHSGSYLKNLSKLSRGRLDLMLCENTALKEKIDKLNKELEKVEPQLT 246
Qy 258 FKNAKLYGPDVGOPRR-KTAKMLKSLFKAGEV-----IDSVTHHHYLYNGRT 304
Db 247 FLRSK---NSIEKPRNREKFLKKFLAMOEKIYLRKRLQIRKIPNKYSDRSLSNST 303
Qy 305 ATREDPLNPDVLD---IFISSQVQFQVVESTPRGKKVWLGETSSAYGGGAPLLSDTFAA 361
Db 304 PKSQDNWTTTQVTPSSSLGVSESVKVLQL-----KQVQVDITE----- 340
Qy 362 GFMWLDKLGLSARMGTVEVNRQVFFGAGN-----YHLVDENFDPLPDYWLSSLFKKLVGTK 417
Db 341 -LVKIPKPNPFSEKLTISNVRNYLNVPGSLDQLPSLTNEFN---VHWNSTVYQELLNLK 395
Qy 418 VLMASSVQGSKRKK 430
Db 396 SNNSSVDGVKTRR 408

RESULT 10
S74760
hypothetical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)
A/Species: *Synechocystis* sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: S74760
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S74760
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-411 <KAN>
A/Cross-references: UNIPROT:P72895; UNIPARC:UPI000000C0C3B; EMBL:D90901; GB:AB001339; NID
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

```

Qy 416 TKVLMAVSV 423
Db 401 LKSLHQI 408

RESULT 11
E91031
probable outer membrane protein EC93221 [imported] - Escherichia coli (strain O157:H7, st
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 05-Oct-2004
C:Accession: E91031
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91031
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-879 <HAY>
A:Cross-references: UNIPROT:Q8XCP4; UNIPARC:UPI0000D0453; GB:BA000007; PIDN:BAB36644.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC93221
C:Superfamily: outer membrane usher protein fimD

Query Match 3.7%; Score 105.5; DB 2; Length 879;
Best Local Similarity 19.9%; Pred. No. 13;
Matches 129; Conservative 69; Mismatches 208; Indels 241; Gaps 33;

Qy 52 VPSFSLSVTIDANLATDPRFLILIGSPKLTARGLSPAYLRFGGTKTDFLIFDPKKEST 111
Db 20 MSGSYVNAWAENEIQFSRFLKGDTKI-DLKRPSSQGYVBPQ--KYNLQVLNKPQPLT 76

Qy 112 PEERSYQSQVNQDICKYSGTPPDVEEKL-----RLEWYQSQLLRHHYQKFKNS 163
Db 77 EYDIYVASENDASKTYACLTPELVAQGLKEDVAKLQWHDGKCKPGQLE----- 130

Qy 164 TYSRSSVDVLYTFANCGLDIFGL-NALLRTADLQMNSSN-----AQLLLDYC----- 211
Db 131 -----GIDIK---ADLSQSAVLSLPQAYLEYTDINWPPSRWDDGISGLIADYSITAGT 182

Qy 212 -----SSKGNYI-SWEL-GNEPNSFL---KKADIFINGSQ----- 241
Db 183 RHEENGCGDSNEISGNGTVGVNLGAWRLRADWQTDYLSHKSNDDDVINGDDTQKNWESR 242

Qy 242 -----IGEDFTQLHKLKSTF-----KNAKLYGDPV 268
Db 243 YAWRALPSLKAKGLIGEDY-----LNSDIFDGFNVYVSGSISTDDQMLPPLRGVAPDI 296

Qy 269 GQPRRTAKMLKSLKAGGEVI-----DSVTHHHYLLNGRTRATREDFLN 312
Db 297 SGVAHTTAKVTVSQ---GRVIYETQVPAGFRIQDLGDSV-----SGTLHRIEQQN 346

Qy 313 PDVLIDPISSVQKVFQVVESTRPK---KWL-----GETS----- 345
Db 347 GQVQEQYDINTASMPF---LTRPGQVRYKLMWGRPQEWGHVHGVEGFFSGGEASWGIANGW 402

Qy 346 SAYGGGAPLLSD-----TFAA-----GFMWLDKL-----GLSAR 374
Db 403 SLYGGA---LADEHYQSAALGVGRDLSVFGGAVAFDITHSHTRLDKETAYGKSLDGNSPR 459

Qy 375 MGI-----EVMVRQVFCAAGNYHLVDENFDPLPDVWLSILLFKKLAVGT---KVLMAVQGS 426
Db 460 LSVSKDFDELNSRVTFAG---YRFEENFTMTSEY-LDASDSEMVTRTGNCKEMYTATYNO 515

Qy 427 KRRKRLVYLHCTNDNRYKEGDTLYAI-----NLHNVTK----- 462
Db 516 NFRDAGVSVLYNTRHYTWDEDEQTNVNMVLSHVFNLGSIKRNMSISMTGYRYEYDQADK 575

Qy 463 --YLRLPYPSFNKQVDKYLRPLRPLGPHGLLSKSVQLNGLTLLKVVDDQT 507
Db 576 GVYISLWMPGSDSTISY---NGNYGSGSDSSQVG---YFSRVDDAT 616

```

RESULT 12

S00652
phosphoribosylamine-glycine ligase (EC 6.3.4.13) - fission yeast (Schizosaccharomyces pombe)
N:Alternate names: AIRase; aminimidazole ribotide synthetase; GARSase; glycineamide rib
N:Contains: phosphoribosylamine-glycine ligase (EC 6.3.4.13); phosphoribosylformylglycin
C:Species: Schizosaccharomyces pombe
C:Date: 07-Sep-1990 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C:Accession: S00652; T40496; T40422
R:McKenzie, R.; Schuchert, P.; Kilbey, B.
Curr. Genet. 12, 591-597, 1987

A:Title: Sequence of the bifunctional adel gene in the purine biosynthetic pathway of th
A:Reference number: S00652; MUID:89003164; PMID:3502942
A:Accession: S00652
A:Molecule type: DNA
A:Residues: 1-788 <MCK>

A:Cross-references: UNIPROT:P20772; UNIPARC:UPI0000132A3F; EMBL:X06601; NID:94903; PIDN:
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21910
A:Accession: T40496
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-788 <WOO>

A:Cross-references: UNIPARC:UPI0000132A3F; EMBL:AL021730; PIDN:CAAL6823.1; GSPDB:GN00067
A:Experimental source: strain 972h-; cosmid c4C3
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21928
A:Accession: T40422
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 604-788 <SEB>

A:Cross-references: UNIPARC:UPI000016908D; EMBL:AL035655; PIDN:CAB38600.1; GSPDB:GN00067
A:Experimental source: strain 972h-; cosmid c405
C:Genetics:
A:Gene: ADEL1; SPDB:SPBC405.01
A:Map position: 2
A:Superfamily: Saccharomyces cerevisiae ADE5 multifunctional protein; phosphoribosylamin

C:Keywords: cyclo-ligase; purine nucleotide biosynthesis
F:5-425/Domain: phosphoribosylamine-glycine ligase homology <PGL>
F:439-767/Domain: phosphoribosylformylglycinamide cyclo-ligase homology <PFCL>
Query Match 3.7%; Score 104.5; DB 1; Length 788;
Best Local Similarity 27.7%; Pred. No. 13;
Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;

QY 297 HYVINGRTATRE--DFLNPDV-LDIFISSVQKVFQVVEST-RPGKKVWLGETSSAY---- 348
DB 424 HIALNPKRTREILTYENSGVSDNGNEFQRIKDLVKSTRPGADADIGGFGGIFDLKQ 483

QY 349 -GGGAPLL-SDTPAAGFMWLDKLGLSAR--MGIEVNRQVFFGAGNYHLVDENFDPL--P 402

DB 484 AGWNPDLVSATDGVGSKLLIALSLNKHDTVGIDLVAMN-----NDLVVQGAEPILFL 537

QY 403 DWYLSLLFKLVGCTKVMASVQSGSKRKLRYVILHCTNDNPRYKEGDLTYALNHNVTK 462

DB 538 DYFATGSLDLKVTSTFVGVGKQAGCALVGGSETSEMPGLYHDGHDYDANGTSVGAISR 597

QY 463 YLRLPYFNSKNQVDKYLRLPLGPHGLLSKSVQLNGLTL--KMVD----DQTLPLPMEKPL 516

DB 598 DDLPLPESFGKDILL-----GLASDGVHNSGSLVRKIVEYSDLEYSVCPCWDKNV 650

QY 517 RPSGSLGLPAFSY 529

DB 651 RLGDSLLIPRIY 663

RESULT 13

F70411
adenylosuccinate synthetase - Aquifex aeolicus

C:Species: Aquifex aeolicus
Query Match 3.7%; Score 104; DB 2; Length 796;
Best Local Similarity 21.6%; Pred. No. 15;

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: F70411
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove
V.

Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: F70411

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-432 <AOF>
A:Cross-references: UNIPROT:O67321; UNIPARC:UPI00000565A2; GB:AE000733; NID:g2983720; PII

A:Experimental source: strain VF5
C:Genetics:
A:Gene: purA
C:Superfamily: adenylosuccinate synthase

Query Match 3.7%; Score 104; DB 2; Length 432;
Best Local Similarity 23.9%; Pred. No. 5.9;
Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;

QY 15 LLLGLGLPLSPGALPRPAQADVVDLD-----PFTQEPHLVSPS 55

DB 51 ILHLLPTGILHEHVKGVIAGM-VVDLEVLHKEVNLEEKGIYVKERLIFSDRAHLVMPY 109

QY 56 FLSVTIDANLATDPRFLILLGSPK--LRTLARGLSPAYL-RFGGTTKTDLFIDPKKESTF 112

DB 110 H-----KLLDSLFEKKKGIGTTLRGIPAYMPKYG--RKGIIRISDLKDKRF 154

QY 113 EERSYQSQVNQDICKYGSIPPDVEEK-----LRLEWYQEOQLLREHYQKPKPNSTY 165

DB 155 ----YTLLEDNDFVK-----NICEKVFCEKFDLDINQIYEEQL----RYFEFPKENV- 199

QY 166 SRSQVDVLYTFANCGLDLIFGLNALLRTADL----QWNSNAQILLDYCSSKGYNISWE 221

DB 200 ----VDLLRFFNTQKGSVLFEAGQGTLLDMDMGTYPPYVTTSSNASAL-----GLSNG 246

QY 222 LGNEPNSFLKKADIFING-----SQL-GEDFIQLHKLRLKSTFFKNAKLYG 265

DB 247 TGMPPKYF---SDAFLGVAKAYTTRVGEPPFPTELKGEGEKREL-----GGYGG 295

QY 266 PDVQGPQR---KTAKMLKSLKAGGEVIDSVTHHHYLLNGRTATREDPLNP----- 313

DB 296 STTGRPRRCGWLDLVALKAYQVNG-----LDGFVITKLDVLDTFDEVKVCVA 343

QY 314 -----DVLDIFISSVQKVFQV--VESTRPCKKXWLGETSSA 347

DB 344 YELDGEVIDYFPASYSSELIRKVPYKTLKG---WKKSTKGA 381

RESULT 14

D97065
transketolase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97065

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97065

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-796 <KUR>

A:Cross-references: UNIPROT:Q97JE3; UNIPARC:UPI0000131969; GB:AE001437; PIDN:AAK79311.1;
A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:
A:Gene: CAC1343
C:Superfamily: phosphoketolase

Query Match 3.7%; Score 104; DB 2; Length 796;
Best Local Similarity 21.6%; Pred. No. 15;

Matches	77; Conservative	51; Mismatches	114; Indels	114; Gaps	18;
Qy	27	GALPRPAQADVVDLDFTEQEPHLVSPFSLVTIDA---NLATDPRF-LILLGSPKLRT 82			
Db	238	GWKPYFVEGEDPETMHKMAETLTDIVTEELINQKARENNDCSRKPWPMIVLRTPK--- 294			
Qy	83	LARGSLPAYLRFQGTKYDFLIFDPKKESTFEERSYQSQVNQDICKYGSIPDPVEEKRL 142			
Db	295	-----GWTGPKFV-----DGVNPEGFRAHQVPLAVDRYHTENLDQLE----- 332			
Qy	143	EW--PYQEOILLRHYQ--KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQ 198			
Db	333	EWLKSYPDELFDENYRLIPELEBLTPGKNRMAANLHAN--GGL-----LLRELTPDPR 386			
Qy	199	WNSNAQLLDYCSSKGYNISWELGNPNPFLKKADIFINGSQIGEDFIQLHKLRL----- 254			
Db	387	-----DYA-----VDVPTPGSTVKQDMIELGKYVRDVVK 415			
Qy	255	-KSTFKNAKLYGPD-----VQPRRTAKMLK---SFLKAGEVIDSVTWHH-- 297			
Db	416	LNEDTRNFRIFGDETWSNRLMAVFEGTKRWLSEIKENPDEFLSNDGRIVDSMLSEHLC 475			
Qy	298	-----YLLNGRTATREDFLNPDVLDIFISSQKVFQVWES--TRPGKKVWLGETS 345			
Db	476	EGWLEGYLLTGRHG-----FFASYEAFRLIVDSMITQHGK--WLKVTS 516			
RESULT 15					
A11489					
probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocua					
C;Species: Listeria innocua					
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004					
C;Accession: A11489					
.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker					
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.					
D.; Jones, L.M.; Karet, U.					
Science 294, 849-852, 2001					
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma					
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,					
A;Title: Comparative genomics of Listeria species.					
A;Reference number: AB1077; MUID:21537279; PMID:11679669					
A;Accession: A11489					
A;Status: preliminary					
A;Molecule type: DNA					
A;Residues: 1-2013 <GLA>					
A;Cross-references: UNIPROT:Q92EK2; UNIPARC:UPI00000CC237; GB:AL592022; PIDN:CAC95689.1;					
A;Experimental source: strain Clip11262					
C;Genetics:					
A;Gene: lin0457					
Query Match 3.7%; Score 104; DB 2; Length 2013;					
Best Local Similarity 20.6%; Pred. No. 60;					
Matches	95; Conservative	64; Mismatches	161; Indels	142; Gaps	21;
Qy	5	SKPALPPPLMLLLGLPLSPGALPRPAQADVVDLD-----FFT 45			
Db	256	AKGPGVPMNLKATL---SGENSAGATYTPAEKTTVNLEENSSNLDSPITAGDNSWAFSM 312			
Qy	46	QEPHLVSPSFLSV---TIDANLATDPRFLILLGSPKLRTLARG-----LSPAYLR 93			
Db	313	KELAFSLKPGGYTTIQWPEIQKKSSENKSFKNL----KLEFLKENGDDIISVNTADPYVIR 368			
Qy	94	FG-----GKTDFLIFDPKKESTFEERSYQSQVNQDICKYGSIPDPVEEKRL 143			
Db	369	FGEPYWSQLSTVNGKANVLVNDDEKQ-----VVEYGPINANIYQRIQVS 412			
Qy	144	WP-----YQQLLRREHYQKFKNSTYSRSSVDVLYTFANCSGLDL 184			
Db	413	MAAKIPADAVKGTGYTGVNVYDSDILVTSTIKITEVTDSTATSDSKVSKTSISEGDV 472			
Qy	185	IFGLNALLRTADLQWN-----SSNA-----QLLLDYCSSKGYNI--SWELGNPNPFLKKAD 234			
Db	473	-----LEWGFMPRISSAAPGVNDLEIVAPIPKIKVLVIYIPNNNSMASMKKLE 520			

Qy	235	IFING-----SOLGEDFIQLHKLRL-----KSTFKNAKLYGPDVGQPRRKTAKWLKS 281			
Db	521	YYQNGKWYSMAPOQTSSGWDFFSKIDQSVNRIEKLKLTSRDGIINDKMDPPYTHGTHRMQNT 580			
Qy	282	FLKAGGEVI---DSVTWHHYLLNGRTATREDFLNPDVLDIFISSVQKVFQVWVEST-RPGK 337			
Db	581	GVKAGESFTLOPESIT---YTDSDKTS-----KAIDTTASSYEKKVQVVEKTSTPAK 629			
Qy	338	---KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMG 376			
Db	630	INGDVFLSSTAGIYKG--FESTIFFNG---DKIAQSVRLG 665			

Search completed: June 5, 2006, 12:21:41
Job time : 21.0785 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 134.728 Seconds
(without alignments)
3728.138 Million cell updates/sec

Title: US-10-645-659A-4
Perfect score: 2841
Sequence: 1 MLRSKPALPPPLMLLLGP.....LPASYSFFVIRNAKVAACI 543

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 7.2.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	543	1	HPSE HUMAN
2	2285	80.4	545	1	HPSE BOVIN
3	2276	80.1	574	2	Q333X9 GRODE
4	2275	80.1	574	2	Q333X8 SPALAX GOLA
5	2263	79.7	574	2	Q333X7 GRODE
6	2262	79.6	574	2	Q333X6 SPALAX JUDA
7	2170	76.4	535	1	HPSE MOUSE
8	2154	75.8	558	2	Q333X5 SPALAX JUDA
9	2138	75.3	536	1	HPSE RAT
10	1648.5	58.0	523	1	HPSE CHICK
11	1320	46.5	533	2	Q4SIF6 TETNG
12	1150.5	40.5	592	1	HPSE2 HUMAN
13	1150.5	40.5	592	2	Q2M1H9 HUMAN
14	1036.5	36.5	597	2	Q4TB80 TETNG
15	742.5	26.1	255	2	Q4TGC8 TETNG
16	699	24.6	515	2	Q8T108 BOMMO
17	417	14.7	543	1	HPSE1 ARATH
18	400	14.1	559	2	Q89F99 BRAJA
19	388	13.7	526	2	Q5SNA6 ORYSA
20	379	13.3	541	2	Q69I16 ORYSA
21	377	13.3	527	2	Q91RC8 SCUBA
22	368	13.0	536	1	HPSE3 ARATH
23	365	12.8	537	2	Q70YJ3 HORVU
24	359.5	12.7	539	2	Q2QN56 ORYSA
25	354	12.4	401	2	Q30324 ARATH
26	353.5	12.4	539	1	HPSE2 ARATH
27	350.5	12.3	529	2	Q6ZUE2 ORYSA
28	323.5	11.4	516	2	Q447R5 SOLUS
29	292	10.3	537	2	Q43S03 SOLUS
30	276.5	9.7	506	2	Q37Q70 SPHAR
31	245.5	8.6	382	2	Q3E8P7 ARATH

32	156	5.5	935	2	Q9VE79 DROME	Q9ve79 drosophila
33	141	5.0	559	2	Q7SFB0 NEUCR	Q7sfb0 neurospora
34	140.5	4.9	536	2	Q2UDS9 ASPOR	Q2uds9 aspergillus
35	138.5	4.9	463	2	Q63T97 BURPS	Q63t97 burkholderi
36	138.5	4.9	670	2	Q3JTG0 BURP1	Q3jtg0 burkholderi
37	138	4.9	1128	2	Q5TT65 ANOGA	Q5tt65 anopheles g
38	136.5	4.8	795	2	Q2ZPT8 SHEPU	Q2zpt8 shewanella
39	135.5	4.8	795	2	Q2X712 GCAMM	Q2x712 shewanella
40	134	4.7	510	2	Q2U0T3 ASPOR	Q2u0t3 aspergillus
41	126.5	4.5	493	2	Q2HK01 THEAC	Q2hk01 thermoplasma
42	125.5	4.4	665	2	Q5S1C3 CRYNE	Q5s1c3 cryptosporidium
43	125.5	4.4	665	2	Q5K7V9 CRYNE	Q5k7v9 cryptosporidium
44	124	4.4	634	2	Q5N1L7 9BACT	Q5n1l7 uncultured
45	124	4.4	765	2	Q4P0C9 USTMA	Q4p0c9 ustilago ma

ALIGNMENTS

RESULT 1
HPSE HUMAN
ID HPSE_HUMAN STANDARD; PRT; 543 AA.
AC Q9Y251; Q53GE5; Q9UL39;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1999, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Heparanase precursor (EC 3.2.2.-) (Heparanase-1) (Hpal) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPR1, HPSE1, HSE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and functional expression of a human heparanase gene.";
RL Blochem. Biophys. Res. Commun. 261:183-187(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and expression.";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
RX PubMed=10395325; DOI=10.1038/10518;
RA Vlodaysky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Becker I.;
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";
RL Nat. Med. 5:793-802(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
RN [5]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE

RP SPECIFICITY.
 RX TISSUE=Placenta;
 RA MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
 RC Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
 RA "Heparanase expression in invasive trophoblasts and acute vascular
 RT damage.";
 RL Glycobiology 10:467-475 (2000).
 RN [16]
 RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
 RT "Molecular properties and involvement of heparanase in cancer
 RT progression and mammary gland morphogenesis.";
 RL J. Mammary Gland Biol. Neoplasia 6:311-322 (2001).
 RN [7]
 RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
 RC TISSUE=Placenta;
 RX PubMed=12713442; DOI=10.1042/BJ20030318;
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
 RA Turner P., Stamps A., McWilliam D., Saville G., Ng S., Mason S.,
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
 RA Farekh R., Stuberfield C.;
 RT "Biochemical characterization of the active heterodimer form of human
 RT heparanase (Hpa1) protein expressed in insect cells.";
 RL Biochem. J. 373:423-435 (2003).
 RN [8]
 RN NUCLEOTIDE SEQUENCE [MRNA].
 RP Pinal M.A., Smedo P.;
 RA "Cloned heparanase from MCF-7 cells.";
 RT Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Small intestine;
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
 RA Tanaka A., Yokoyama S.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC TISSUE=Pancreas;
 RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
 RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien P.,
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [11]
 RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
 RX PubMed=11123890; DOI=10.1021/bi002080p;
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
 RA Gready J.E., Parish C.R.;
 RT "Identification of active-site residues of the pro-metastatic
 RT endoglycosidase heparanase.";
 RL Biochemistry 39:15659-15667 (2000).
 RN [12]
 RN N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
 RP ASN-217; ASN-238 AND ASN-459.

RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
 RA Shimizu S., Ishida K., Wierzbicka M.K., Osada H.;
 RT "Secretion of heparanase protein is regulated by glycosylation in
 RT human tumor cell lines.";
 RL J. Biol. Chem. 279:2697-2703 (2004).
 RN [13]
 RP SUBCELLULAR LOCATION.
 RX PubMed=15292202; DOI=10.1074/jbc.M402131200;
 RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
 RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
 RT "Heparanase uptake is mediated by cell membrane heparan sulfate
 RT proteoglycans.";
 RL J. Biol. Chem. 279:44084-44092 (2004).
 RN [14]
 RN BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
 RP PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
 RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
 RT "Heparanase processing by lysosomal/endosomal protein preparation.";
 RL FEBS Lett. 579:2334-2338 (2005).
 RN [15]
 RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
 RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
 RA Abboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
 RA Elgavish S., Peretz T., Vlodavsky I.;
 RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
 RT human proheparanase.";
 RL J. Biol. Chem. 280:13568-13575 (2005).
 RN [16]
 RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
 RX PubMed=15760902; DOI=10.1074/jbc.M414546200;
 RA Levy-Adam F., Abboud-Jarrous G., Guerrini M., Beccati D.,
 RA Vlodavsky I., Ilan N.;
 RT "Identification and characterization of heparin/heparan sulfate
 RT binding domains of the endoglycosidase heparanase.";
 RL J. Biol. Chem. 280:20457-20466 (2005).
 RN [17]
 RP VARIANT SER-260.
 RX PubMed=15334672;
 RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
 RA Shi L.H., Shen H.W., Xu B.;
 RT "Heparanase mRNA expression and point mutation in hepatocellular
 RT carcinoma.";
 RL World J. Gastroenterol. 10:2795-2799 (2004).
 CC -I- FUNCTION: Endoglycosidase which is a cell surface and
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core
 CC proteoglycans. Also implicated in the extravasation of leukocytes
 CC and tumor cell lines. Due to its contribution to metastasis and
 CC angiogenesis, it is considered to be a potential target for anti-
 CC cancer therapies.
 CC -I- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
 CC lower extent, by heparin and sulfamin and activated by calcium and
 CC magnesium (by similarity).
 CC -I- BIOPHYSICOCHEMICAL PROPERTIES:
 CC pH dependence:
 CC Optimum pH is 4-6;
 CC -I- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
 CC subunits, the proteolytic products.
 CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
 CC Secreted, internalised and transferred to late endosomes/lysosomes
 CC as a proheparanase. In lysosomes, it is processed into the active
 CC form, the heparanase. The uptake or internalisation of
 CC proheparanase is mediated by HSPGs. Heparin appears to be a
 CC competitor and retain proheparanase in the extracellular medium.
 CC -I- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
 CC weakly expressed in lymph node, thymus, peripheral blood
 CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor
 CC tissues.
 CC -I- PTM: Proteolytically processed. The cleavage of the 65 kDa form
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
 CC product. The active form, the 8/50 kDa heterodimer, is resistant
 CC to degradation. Complete removal of the linker peptide appears to
 CC be a prerequisite to the complete activation of the enzyme.

```
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.

Query Match 100.0%; Score 2841; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 2e-205;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFDFTQPLHLVSPFLSVT 60
DB 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFDFTQPLHLVSPFLSVT 60
QY 61 IDANLATDPRFLLGLSPKLTLAGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLLGLSPKLTLAGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKRLLEWYQEQLLREHYQKFKNSTYSSRSSVDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDPVEEKRLLEWYQEQLLREHYQKFKNSTYSSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALRTADLOWNSSNAQLLDYCSSKGYNISWELGNERNPSFLKKADIFINGS 240
DB 181 GLDLIFGLNALRTADLOWNSSNAQLLDYCSSKGYNISWELGNERNPSFLKKADIFINGS 240
QY 241 QLCGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKRTAKMLKSFLLKAGGEVIDSVTWHYYL 300
DB 241 QLCGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKRTAKMLKSFLLKAGGEVIDSVTWHYYL 300
QY 301 NGRTATREDFLNPVDLIDIFISSVQKVFQVVESTPRGKVKWLGTSSAYGCGAPLLSDTFA 360
DB 301 NGRTATREDFLNPVDLIDIFISSVQKVFQVVESTPRGKVKWLGTSSAYGCGAPLLSDTFA 360
QY 361 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDLPDYWLSSLFPKLVGTKVL 420
DB 361 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDLPDYWLSSLFPKLVGTKVL 420
QY 421 ASVQGSKRKRLRYLHCTNTDNPYKEGDLTLVAINLHNVTYKLRLPYPSPNKQVDKYL 480
DB 421 ASVQGSKRKRLRYLHCTNTDNPYKEGDLTLVAINLHNVTYKLRLPYPSPNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKMVDQDTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVQLNGLTLKMVDQDTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540
QY 541 ACI 543
DB 541 ACI 543

RESULT 2
ID HPSE BOVIN STANDARD; PRT; 545 AA.
AC Q9MY0;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
GN Name:HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;
RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
```

```
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC
CC EMBL: AF281160; AAF87301.2; -; mRNA.
CC InterPro: IPR005199; Glyco_hydro_79_N.
CC Pfam: PF03662; Glyco_Hydro_79n; 1.
CC KEGG: Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
CC Signal.
CC FT SIGNAL 1 37 By similarity.
CC FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
CC FT PROPEP 112 159 /FTid=PRO_0000042256.
CC FT CHAIN 160 545 /FTid=PRO_0000042257.
CC FT CHAIN 160 545 Heparanase 50 kDa subunit (By
CC similarity).
CC FT /FTid=PRO_0000042258.
CC FT REGION 160 164 Heparin/HS-binding (Potential).
CC FT REGION 272 282 Heparin/HS-binding (Potential).
CC FT ACT_SITE 227 227 Proton donor (Potential).
CC FT ACT_SITE 345 345 Nucleophile (Potential).
CC FT CARBOHYD 164 164 N-linked (GLCNAC...) (Potential).
CC FT CARBOHYD 219 219 N-linked (GLCNAC...) (Potential).
CC FT CARBOHYD 461 461 N-linked (GLCNAC...) (Potential).
CC SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD85B933 CRC64;

Query Match 80.4%; Score 2285; DB 1; Length 545;
Best Local Similarity 80.0%; Pred. No. 1.9e-163;
Matches 436; Conservative 34; Mismatches 73; Indels 2; Gaps 1;

QY 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFDFTQPLHLVSPFLS 58
DB 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFDFTQPLHLVSPFLS 60
QY 59 VTIDANLATDPRFLLGLSPKLTLAGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 118
DB 61 FTIDANLATDPRFLLGLSPKLTLAGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
QY 119 QSVNQDICKYGSIPDPVEEKRLLEWYQEQLLREHYQKFKNSTYSSRSSVDVLYTFAN 178
DB 121 LSQSNQDICKYGSIPDPVEEKRLLEWYQEQLLREHYQKFKNSTYSSRSSVDVLYTFAN 180
QY 179 CSGDLIFGLNALRTADLOWNSSNAQLLDYCSSKGYNISWELGNERNPSFLKKADIFIN 238
DB 181 CSGDLIFGLNALRTADLOWNSSNAQLLDYCSSKGYNISWELGNERNPSFLKKADIFIN 240
QY 239 GSQLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKRTAKMLKSFLLKAGGEVIDSVTWHYY 298
```



```
QY 60 TIDANLATDPRFLILLGSPKRLTLARGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQ 119
DB 91 TIDANLATDPRFLTLGSPKRLALARGLSPAYLRFGGTKTDFLI FDPKKEPSHEERSYK 150
QY 120 SQVNQDICKYGSIPPDVEEKRLRLEWYQEQQLLLREHYQKFKNSTYSSRSVDVLYTFANC 179
DB 151 SQVNHDICRSGAIPAVVVRRLQVEWPFQEQQLLLREQYQKEFKNSTYSSRSVDMLYTFARC 210
QY 180 SGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKAIDIFING 239
DB 211 SGLDLIFGLNALLRTADFRWNSSNAQLLLNYCSSKNYDISWELGNEPNSFWKKAHISIDG 270
QY 240 SOLGEDFIQLHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 299
DB 271 LQLEDYIELHKLKRSKSTLKNVKLYGPDVGOPRGKTVKLLRSFLKAGGEVIDSVTWHYY 330
QY 300 LNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKVKWLGETSSAYGGGAPLLSDTF 359
DB 331 LNGRIATKEDFLSPDVLDTFILSVQKILQVVESTRPGKVKWLGETSSAYGGGAPLLSNTF 390
QY 360 AAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 419
DB 391 AAGFMWLDKLGLSQMGIEVVMRQVFFGAGNYHLVDKNFELPDYWLSSLFLFKLVGSKVL 450
QY 420 MASVQSGSKRKLRYLHCTNTDNPRYKEGDLTLAYALNHNVTYKYLRLPYPPSNKQVDKYL 479
DB 451 MARVKGDRSKRLRYLHCTNINHPRYQEGDLTLAYALNLYNVTYKYLRLPYQLFNKPVDKYL 510
QY 480 LRPLGPHGLLSKSVQNLGLTLKMVDDQTLPLPMEKPLRPGSSGLPAPFSYFFVIRNAKV 539
DB 511 VKPLGPGGLLSKSVQNLGQALKWDDQTLPALTEKPLRPGSSGLPAPFSYGFVIRNAKV 570
QY 540 AACI 543
DB 571 AACL 574
```

```
RESULT 5
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
DS Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FD19DCBBD99DE CRC64;
```

```
Query Match 79.7%; Score 2263; DB 2; Length 574;
Best Local Similarity 80.1%; Pred. No. 9.4e-162;
Matches 436; Conservative 42; Mismatches 60; Indels 6; Gaps 2;
QY 4 RSKPALPPPLMLLLL-----GPLGLSPGALPRPAQADVVLDLFFQEPHLHVSPLSV 59
DB 33 RCQGPPEMLRLSLLLMLWGLSPLVQCIL--AAQADVVVELBFSTQRLHLVSPSLSI 90
```

```
QY 60 TIDANLATDPRFLILLGSPKRLTLARGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQ 119
DB 91 TIDANLATDPRFLTLGSPKRLALARGLSPAYLRFGGTKTDFLI FDPKKEPSHEERSYK 150
QY 120 SQVNQDICKYGSIPPDVEEKRLRLEWYQEQQLLLREHYQKFKNSTYSSRSVDVLYTFANC 179
DB 151 SQVNHDICRSGAIPAVVVRRLQVEWPFQEQQLLLREQYQKEFKNSTYSSRSVDMLYTFARC 210
QY 180 SGLDLIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNISWELGNEPNSFLKKAIDIFING 239
DB 211 SGLDLIFGLNALLRTADFRWNSSNAQLLLNYCSSKNYDISWELGNEPNSFWKKAHISIDG 270
QY 240 SOLGEDFIQLHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 299
DB 271 LQLEDYIELHKLKRSKSTLKNVKLYGPDVGOPRGKTVKLLRSFLKAGGEVIDSVTWHYY 330
QY 300 LNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKVKWLGETSSAYGGGAPLLSDTF 359
DB 331 LNGRIATKEDFLSPDVLDTFILSVQKILQVVESTRPGKVKWLGETSSAYGGGAPLLSNTF 390
QY 360 AAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 419
DB 391 AAGFMWLDKLGLSQMGIEVVMRQVFFGAGNYHLVDKNFELPDYWLSSLFLFKLVGSKVL 450
QY 420 MASVQSGSKRKLRYLHCTNTDNPRYKEGDLTLAYALNHNVTYKYLRLPYPPSNKQVDKYL 479
DB 451 MARVKGDRSKRLRYLHCTNINHPRYQEGDLTLAYALNLYNVTYKYLRLPYQLFNKPVDKYL 510
QY 480 LRPLGPHGLLSKSVQNLGLTLKMVDDQTLPLPMEKPLRPGSSGLPAPFSYFFVIRNAKV 539
DB 511 VKPLGPGGLLSKSVQNLGQALKWDDQTLPALTEKPLRPGSSGLPAPFSYGFVIRNAKV 570
QY 540 AACI 543
DB 571 AACL 574
```

```
RESULT 6
Q333X6_SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
DS Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;
```



```
QY 180 SGLDIFGLNALLRTADLOWNSSNAQLLDYCSSKGVNLSWELGNBPNPSFLKKADIFING 239
DB 211 SGLDIFGLNALLRTADFRWNSSNAQLLDYCSSKGVNLSWELGNBPNPSFLKKAHISIDG 270
QY 240 SOLGEDFIOLKLLRSTFKNAKLYGPDVQGPQRRKTAQMLKSLFKAGGEVIDSVTWHYY 299
DB 271 LQLEDYIELKLLKSTLKNVLYGPDVQGPQPKTVKLLRS-----YY 314
QY 300 LNGRTATREDFLNPVDLIFISSVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTF 359
DB 315 LNGRTATREDFLNPVDLIFISSVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSNTF 374
QY 360 AAGFWMLDKGLSARNGIEVWVRQVFFGAGNYHLDVNDPDLVDYWLISLLFKLVGCKVL 419
DB 375 AAGFWMLDKGLSARNGIEVWVRQVFFGAGNYHLDVNDPDLVDYWLISLLFKLVGSKVL 434
QY 420 MASVQSKRKLRYVLLHCTNTDNPRIKGGDLTYALNHNVTYKYLRYPFPSNKQVDKYL 479
DB 435 MARVQGPDRSKLRYVLLHCTNINHPRIQGGDLTYALNHNVTYKYLRYPFPSNKQVDKYL 494
QY 480 LRPLGPHGLSKSVQLNGLTLMKVDQDTLPPLMEKPLRPGSSGLPFAFSYFFVIRNAKV 539
DB 495 VIPLGPGGLSKSVQLNGQAKWVDQDTLPALTEKPLRPGSSGLPFAFSYFFVIRNAKV 554
QY 540 AACI 543
DB 555 AACL 558

RESULT 9
HPSE RAT STANDARD; PRT; 536 AA.
AC Q71RP1; Q90ZF8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).
GN Name=Hase; Synonyms=Hep;
OS Rattus norvegicus (Rat);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line.";
RL J. Biol. Chem. 277:32459-32465(2002).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
```

```
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL: AF395908; AA015189.1; -; mRNA.
CC EMBL: AF184967; AA04563.1; -; mRNA.
CC RGD: 61969; HpsE.
CC InterPro: IPR005199; Glyco_hydro_79_N.
CC Pfam: PF03662; Glyco_Hydro_79n; 1.
CC KEGG: Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 /FTid=PRO_0000042266.
FT FT Linker peptide (By similarity).
FT FT /FTid=PRO_0000042267.
FT CHAIN 151 536 Heparanase 50 kDa subunit.
FT FT /FTid=PRO_0000042268.
FT REGION 151 155 Heparin/HS-binding (By similarity).
FT REGION 263 273 Heparin/HS-binding (By similarity).
FT ACT_SITE 218 218 Proton donor (Potential).
FT ACT_SITE 336 336 Nucleophile (Potential).
FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 75.3%; Score 2138; DB 1; Length 536;
Best Local Similarity 76.3%; Pred. No. 2.3e-152;
Matches 408; Conservative 49; Mismatches 78; Indels 0; Gaps 0;

QY 9 LPPLMLLLGLPLGPGALPPRAQADVVLDLFFTOEPLHLVSPSPSLVTIDANLATD 68
DB 2 LRPLLLMLWGRLCALTQGTTPAGTAPTQDVVDLEFYTKRLFQSVSPSLTIDASLATD 61
QY 69 PRFLILGSPKLTARGLSPAYLRFGGTKDFLIFDPKKESTFEERSYWSQVNDICK 128
DB 62 PRFLFTPLGSPRLALARGSLPAYLRFGGTKDFLIFDPKKESTFEERSYWSQVNDICK 121
QY 129 YGSIPDPVEKLRLEWYPYQQLLRHVKFKNSTYSRSSVDVLYTPANCSGDLIFGL 188
DB 122 SERVSADVLRLQLNWFQFELLRLREQYQREKNTYSRSSVDVLYTPANCSGDLIFGL 181
QY 189 NALLRTADLOWNSSNAQLLDYCSSKGVNLSWELGNBPNPSFLKKADIFINGSLGEDFTQ 248
DB 182 NALLRTADLOWNSSNAQLLDYCSSKGVNLSWELGNBPNPSFLKKADIFINGSLGEDFTQ 241
QY 249 LHKLLRSTFKNAKLYGPDVQGPQRRKTAQMLKSLFKAGGEVIDSVTWHYYLNGRTATRE 308
DB 242 LHKLLRSTFKNAKLYGPDVQGPQRRKTAQMLKSLFKAGGEVIDSVTWHYYLNGRTATRE 301
QY 309 DFLNPVDLIFISSVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFPAAGFWMLDK 368
DB 302 DFLNPVDLIFISSVQVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFPAAGFWMLDK 361
QY 369 LGLSARNGIEVWVRQVFFGAGNYHLDVNDPDLVDYWLISLLFKLVGCKVLMAVQSKR 428
```



```

Db 362 LGLSAQLGIEVMQVFFGAGNYHLVDENPEPLDYWLSSLFLLKLVGPKULMSRVKGPDR 421
QY 429 RKLRYVYLHCTNTDNPRYKEGDLTYALNHNVTYKRLPYFPFSKNQVDYKLLRLPLGPHGL 488
Db 422 SKRLVYLHCTNVVHPRYREGDLTYLVNLHNVTYKRLPLPPMFMSRPVDKYLKPFSGDGL 481
QY 489 LSKSVQLNGTLKMWDDOTLPLMEKDLRPGSSGLPFAFSYFPFVIRNAKVAACI 543
Db 482 LSKSVQLNGTLLKMWDEQTLPALTEKPLPAGSSLSVPFAFSYGFVIRNAKIAACI 536

RESULT 10
HPSE_CHICK STANDARD; PRT; 523 AA.
AC Q90YK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldshmidt O., Zcharia E., Angorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Fecker I., Mitran E., Vlodavsky I.;
RT "Expression pattern and secretion of human and chicken heparanase are
RT determined by their signal peptide sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC (By similarity).
CC -!- PTM: N-glycosylated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AY037007; AAK82648.1; -; mRNA.
DR Ensembl: ENSGALG0000011203; Gallus gallus.
DR InterPro: IPR005199; Glyco_hydro_79_N.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 523 Heparanase.
FT FTID=PRO_0000042259.
FT REGION 137 141 Heparin/HS-binding (By similarity).
FT REGION 250 260 Heparin/HS-binding (By similarity).
FT ACT_SITE 204 204 Proton donor (Potential).
FT ACT_SITE 323 323 Nucleophile (Potential).
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;
Query Match 58.0%; Score 1648.5; DB 1; Length 523;
Best Local Similarity 60.3%; Pred. No. 2e-115;
Matches 321; Conservative
QY 13 LMILLGLPLGSPALPRQAQDVVDLFFQTEPLHLVSPSLVTYDANLATDREL 72
Db 2 LVLLLLLVLLAVPP-----RRTAELQLGKEPIGAVSPAFSLTLDASLATPRFV 52
QY 73 ILLGSPKRLTLARGLSPAYLRFQGTDTDLIFDPKKESTFEERSYQSQVNQDICKYGI 132

```

```

Db 53 ALLRHPKLTLASGLSPGLRFGGTSTDFLFPNPKDSTWEEKVLSFQA-KDVCENWPS 111
QY 133 PPDVEEKLRLWPYQEOQLLEHYQKFKPNSTYGRSSVDVLYTFANCGLDILFGLNALL 192
Db 112 FAVVPKLLLTQWPLQEKLLLAELHWSKXKNTTITRSTDLILHTFASSSGFRLVFGLNALL 171
QY 193 RTADLQWSSNAQLLDYCCSSKGYNISWELGNEPNSFLKADIFINGSOLGDEFTQLHKL 252
Db 172 RRAGLQWSSNAQLLDYCAQRSYNISWELGNEPNSFRKSGICIDGFLGRHDFVHLRQL 231
QY 253 L-RKSTFKNAKLYGPDVGPQRRKTKAKMLKSLKAGGEVIDSVTWHYYLNGRTATREDFL 311
Db 232 LSQPLRYHAELYGLDVGPQPKHTQHLLRSPKSGKKAIDSVTWHYYVNGRSATREDFL 291
QY 312 NPVDLDIFISSVQKVFQVVESTRPQKVKWLGETSSAYGGAPLLSDTFAAGFMWLDKGL 371
Db 292 SPEVLDSFATAIHDVLGIVEATVPGKVKWLGETSSAYGGAPQLSNTYVAGFMWLDKGL 351
QY 372 SARNGIEVVMRQVFFGAGNYHLVDENPEPLDYWLSSLFLLKLVGPKULMASVQSGKRRKL 431
Db 352 AARREGIDVMRQVFFGAGSYHLVDAGFKPLPDYWLSSLFLLKLVGPKULMASVQSGKRRKL 411
QY 432 RVLHCTNTDNPRYKEGDLTYALNHNVTYKRLPYFPFSKNQVDYKLLRLPLGPHGLLSK 491
Db 412 RVLHCTNPRHPKYREGDVTFLFALNLSNVTQSLQPKQLWSKSDVDQYLLPHGKDSILSR 471
QY 492 SVQLNGTLKMWDDOTLPLMEKDLRPGSSGLPFAFSYFPFVIRNAKVAACI 543
Db 472 EVQLNGRLQLQWDDDTLPALEHMAALAPGSTLGLPFAFSYGFVIRNAKIAACI 523

RESULT 11
Q4SYF6 TETNG PRELIMINARY; PRT; 533 AA.
ID Q4SYF6 TETNG PRELIMINARY; PRT; 533 AA.
AC Q4SYF6;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00010356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```


21-FEB-2006, entry version 1.
DE Heparanase 2.
GN Name=HPSE2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RG NIH MGC project;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: BC112356; AAIL2357.1; -; mRNA.
SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;

Query Match 40.5%; Score 1150.5; DB 2; Length 592;
Best Local Similarity 43.4%; Pred. No. 9.2e-78;
Matches 249; Conservative 83; Mismatches 189; Indels 53; Gaps 9;

QY 20 PLGLSPGAL-----PRPA-----QAQDVVDLDFFTQEPHLHVS 55
DB 18 PPACLAGALYLLALLHLSSAGDRPLPDRAGLKEKTLILDVSTKPNVTNEN 77
QY 56 FLSVTTIDANLDPRLILIGSPKRLTLARGLSPAYLRFQGTGTLFLIF----DPKKEST 111
DB 78 FLSLQLDPSIHD-GWLDLFLSSKRLVTLARGLSPAFRLFGCKRTDPLFQFNLRNPAKSRG 136
QY 112 FEERSTWQSQVNODI-----CKGSGTPPDVEEKLRLFWPQEQI-LLREHYQK 158
DB 137 GCGPDYLYKNYEDIVRSVDVALDKQKQKIAQ-HPDVMLELQREKAAQMHVLILKEQFSN 195
QY 159 KFNKSTYSRSSVDVLYTFANCGLDILFGLNALIRTDALQNSNAQLLDYCSKCYNI 218
DB 196 TYSNLLLTARSCLKYFNADCSGLHIFALNALRNPNNSWSSALSLLKYISASKYINI 255
QY 219 SWELGNPNPSFLKKADIFINGSQLGDFIQHLKLLRK-STFFKNAKLYGPDVGOPRRKTA 277
DB 256 SWELGNPNPNRYTHGKAVNGSQLGKDYIQLSLQPIRYSRASLYGPNIGRKNVIA 315
QY 278 MLKSFLLKAGEVIDSVTHHYHLYNGRTATREDFLNPDVLIDFISVQKVFQVVESTPQK 337
DB 316 LLDGFMKVGASTVDAVTWQHCYIDGRVVKVMDFLKTRLLDLSQIRKIQKVNTYTPGK 375
QY 338 KVMILGTSSAYGGAPLLSDTPFAGFMWLDKLGLSARMGLIEVVMRQVFFGAGNVHLVDEN 397

376 KIWLGVVTTTSAGGTNNLSDSYAAGFLWNLTLGLMANQGDVIRHSFFDFHYNHLVDQN 435
QY 398 FDPILPDYWLSSLKPKLVGTVMASVQGSKR-----KLRYVLHCTNTDPRYKEG 448
DB 436 FNLDPDYWJLSLLYKRLIGPKVLAVHVAGLQKRPGRVIRDKLRIYAHCTNHHNNHYVRG 495
QY 449 DLTLYAINLHNVTKYLRLPYPFNSKNQVDKYLRLPLGPHGLLSKSVQLNGLTLRMVDDOTL 508
DB 496 SITFLINLHRSRKIKLTGTLRDKLVHQYLLQPYGQGLKSKSVQLNQGLVMVDDOTL 555
QY 509 PPLMEKPLRPGSSSLGLPAFYSYFVIRNAKVAAC 542
DB 556 PELKPRPLRAGRTLVIPTVTMGFFVKNVNALAC 589

RESULT 14
Q4TB80_TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG0003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Athouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Blemond C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutancan J.-P., Gouzy J.,
RA Parra G., Lardier G., Chappelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolious H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: CAAB01007180; CAP89852.1; -; Genomic_DNA.
DR NON TER 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match 36.5%; Score 1036.5; DB 2; Length 597;
Best Local Similarity 41.1%; Pred. No. 3.8e-69;
Matches 241; Conservative 88; Mismatches 205; Indels 53; Gaps 14;

QY 3 LRSPALPPPLMLLLGLPLGSPCALPRPAQA-----QDVVDLDFFTQEPHLHVS 54
DB 16 LASLAALLVP--LVLSFPYS--SSSTQYRPAVGKRGPGFVETRLILLDNTKSPIRVLMD 71
QY 55 SFLSVTTIDANLDPRLILIGSPKRLTLARGLSPAYLRFQGTGTLDFLDPKKE----- 109

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 106.86 Seconds
(without alignment)
2293.354 Million cell updates/sec

Title: US-10-645-659A-3
Perfect score: 2800
Sequence: 1 MLRPLLLWGLRLALQTG.....VPAFSYGFVIRNAKIAACI 536

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_8.*

1: Geneseqpl980s.*
2: Geneseqpl990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*
10: Geneseqp2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2800	100.0	536	5 ABB07812	Abb07812 Rat hepar
2	2800	100.0	536	8 ADR88209	Adr88209 Rat hepar
3	2800	100.0	536	8 ADT78176	Adt78176 Rat hepar
4	2800	100.0	536	9 ADY27035	Ady27035 Rat hepar
5	2800	100.0	536	9 AEA42425	Aea42425 Rat hepar
6	2594.5	92.7	535	9 ADY27033	Ady27033 Murine he
7	2590.5	92.5	535	3 AAB08851	Aab08851 A murine
8	2590.5	92.5	535	3 ABB07811	Abb07811 Mouse hepar
9	2590.5	92.5	535	7 ADG88834	Adg88834 Mouse hpa
10	2590.5	92.5	535	8 ADL16413	Adl16413 Mouse hepar
11	2590.5	92.5	535	8 ADM48750	Adm48750 Mouse hpa
12	2590.5	92.5	535	8 ADR88208	Adr88208 Mouse hepar
13	2590.5	92.5	535	8 ADT78175	Adt78175 Mouse hepar
14	2590.5	92.5	535	9 AEA42424	Aea42424 Mouse hepar
15	2151	76.8	543	2 ADY27034	Ady27034 Bovine hepar
16	2126	75.9	543	2 AAY17082	Aay17082 Human hepar
17	2126	75.9	543	7 AAB86206	Aab86206 Human hepar
18	2126	75.9	543	7 ADD18950	Adl18950 Human dis
19	2126	75.9	543	8 ADK52086	Adk52086 Human ato
20	2126	75.9	543	8 ADM48759	Adm48759 Human hpa
21	2126	75.9	543	8 ADN05074	Adn05074 Antipsori
22	2126	75.9	543	8 ADN04902	Adn04902 Antipsori
23	2126	75.9	543	8 ADQ80372	Adq80372 Heparanas

24	2126	75.9	543	8 ADR88210	Adr88210 Human pre
25	2126	75.9	543	8 ADP25079	Adp25079 PRO poly
26	2126	75.9	543	8 ADT78177	Adt78177 Human hepar
27	2126	75.9	543	9 ADY27036	Ady27036 Human hepar
28	2126	75.9	543	9 AEA42426	Aea42426 Human hepar
29	2126	75.9	588	2 AAY30124	Aay30124 A human p
30	2125.5	75.9	545	6 ABP56822	Abp56822 Human hepar
31	2125.5	75.9	545	7 ADE16012	Ade16012 G-coupled
32	2125.5	75.9	545	8 ADL93951	Adl93951 Human G-c
33	2123	75.8	543	2 AAY02345	Aay02345 A human h
34	2123	75.8	543	3 AAY57590	Aay57590 Human hepar
35	2123	75.8	543	3 AAB08849	Aab08849 Amino aci
36	2123	75.8	543	3 AAY52990	Aay52990 Human hepar
37	2123	75.8	543	4 AAY97635	Aay97635 Human hepar
38	2123	75.8	543	5 ABB07813	Abb07813 Human hepar
39	2123	75.8	543	7 ADG88800	Adg88800 Human hpa
40	2123	75.8	543	8 ADL16379	Adl16379 Human hepar
41	2123	75.8	543	8 ADM48716	Adm48716 Human hpa
42	2123	75.8	543	9 AEA42466	Aea42466 Human hepar
43	2123	75.8	543	10 AEE96848	Aee96848 Human hepar
44	2123	75.8	592	2 AAY02346	Aay02346 A human h
45	2123	75.8	592	3 AAB08850	Aab08850 Amino aci

ALIGNMENTS

RESULT 1

ABB07812 ID ABB07812 standard; protein; 536 AA.

AC ABB07812;

DT 03-JUL-2002 (first entry)

DE Rat heparanase sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;

KW anti-protozoan; neuroprotective; heparin; rat.

OS Rattus rattus.

XX Key Location/Qualifiers

FT Peptide 1..16 /note= "putative signal peptide"

FT Protein 17..536 /note= "mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2002-338926/37.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful to treat various heparin-related disorders and the signal peptide is useful in production of membrane-targeted or secreted recombinant proteins.

XX Disclosure; Fig 1a; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid, encoding a polypeptide with heparanase catalytic activity. The signal peptide of the nucleic acid can be used to express membrane-associated or secreted proteins in heterologous expression systems. The encoded

CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC invasion, and to intervene with pathologies associated with impaired
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoa and bacterial infections or
CC disintegration of neurodegenerative plaques. The present sequence
CC represents a rat heparanase protein sequence used in similarity studies
XX
SQ Sequence 536 AA;

Query Match 100.0%; Score 2800; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 8.6e-260; Mismatches 0; Indels 0; Gaps 0;
Matches 536; Conservative 0;

QY 1 MLRPLLLMLWGLRALTGTPAGTAKDQVLEFYTKELFQSVSPFLSIITIDASLAT 60
DB 1 MLRPLLLMLWGLRALTGTPAGTAKDQVLEFYTKELFQSVSPFLSIITIDASLAT 60

QY 61 DPRFLFTPLSPLRALARGLSPAYLRGGTKTDFLIPDPNKEPTSEERSYQSDNNNDIC 120
DB 61 DPRFLFTPLSPLRALARGLSPAYLRGGTKTDFLIPDPNKEPTSEERSYQSDNNNDIC 120

QY 121 GSERVSADVLKQMWEPFQELLRLREYQREPKNSTYSSVDMLYSPAKCSRLDLIFG 180
DB 121 GSERVSADVLKQMWEPFQELLRLREYQREPKNSTYSSVDMLYSPAKCSRLDLIFG 180

QY 181 LNALLTPTDLRWNSNAQLLLNCCSSGYNISWELGNEPNSFWKKAQISIDGLQGBDFV 240
DB 181 LNALLTPTDLRWNSNAQLLLNCCSSGYNISWELGNEPNSFWKKAQISIDGLQGBDFV 240

QY 241 ELHKLQKSAFNQAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLNGRVATK 300
DB 241 ELHKLQKSAFNQAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLNGRVATK 300

QY 301 EPLSSDVLDTFLSVQKILKVTKEPMPGKVMWGETSSAYGGAPLLSTPAAGFWMLD 360
DB 301 EPLSSDVLDTFLSVQKILKVTKEPMPGKVMWGETSSAYGGAPLLSTPAAGFWMLD 360

QY 361 KLGLSAQLGIEVMVROVFFGAGNHYLVDFENPEPLPDYWLSSLFKKLGPKVMSRVKGPD 420
DB 361 KLGLSAQLGIEVMVROVFFGAGNHYLVDFENPEPLPDYWLSSLFKKLGPKVMSRVKGPD 420

QY 421 RSKLRVYLCTNVYHRYREGDITLVVNLNHNVTKHLKLPMPFSPVDKYLKLPFGSDG 480
DB 421 RSKLRVYLCTNVYHRYREGDITLVVNLNHNVTKHLKLPMPFSPVDKYLKLPFGSDG 480

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536
DB 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536

RESULT 2
ADR88209
ID ADR88209 standard; protein; 536 AA.
XX
AC ADR88209;
XX
DT 18-NOV-2004 (first entry)
XX
DE Rat heparanase.
XX
KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
KW rat; heparanase; enzyme.
XX
OS Rattus norvegicus.
XX
FH
FT Peptide
FT 1. .16
FT /label= Signal_peptide

FT Protein 17. .536
FT /label= Mature_heparanase
XX US2004170631-A1.
XX
XX 02-SEP-2004.
XX
XX 28-NOV-2003; 2003US-00722502.
XX
XX 02-SEP-1997; 97US-00922170.
XX 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
PR 22-AUG-2003; 2003US-00645659.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VGEL/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI, 2004-625084/60.
XX
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
XX complex.
XX
XX Claim 2; SEQ ID NO 3; 58pp; English.
XX
XX The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an antibody or its portion
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease) for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is rat heparanase.
XX
XX Sequence 536 AA;

```

Query Match      100.0%; Score 2800; DB 8; Length 536;
Best Local Similarity 100.0%; Pred. No. 8.6e-260; Indels 0; Gaps 0;
Matches 536; Conservative 0; Mismatches 0;

QY 1 MLRPLLLWLGRLALQTQGTAGTAPTKDVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
   |||||
Db 1 MLRPLLLWLGRLALQTQGTAGTAPTKDVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
   |||||

QY 61 DPRFLTLGSPRLALAGLSPAYLRFGGTKTDFLIIDPNKEPTSEERSYQSQNDNDIC 120
   |||||
Db 61 DPRFLTLGSPRLALAGLSPAYLRFGGTKTDFLIIDPNKEPTSEERSYQSQNDNDIC 120
   |||||

QY 121 GSERVSADVLRLKQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
   |||||
Db 121 GSERVSADVLRLKQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
   |||||

QY 181 LNALLRTPDLRWNSSNAQLLNTCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   |||||
Db 181 LNALLRTPDLRWNSSNAQLLNTCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   |||||

QY 241 ELHKLQKSAFQAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
   |||||
Db 241 ELHKLQKSAFQAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
   |||||

QY 301 EDFLSSDVLDTFLSVQKILKVTKEWTPGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360
   |||||
Db 301 EDFLSSDVLDTFLSVQKILKVTKEWTPGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360
   |||||

QY 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLKLLPKLGPVKVMSRVKGPD 420
   |||||
Db 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLKLLPKLGPVKVMSRVKGPD 420
   |||||

QY 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNTVKHLKLPMPFSPRVDKYLKPFSGSDG 480
   |||||
Db 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNTVKHLKLPMPFSPRVDKYLKPFSGSDG 480
   |||||

QY 481 LLSKSVQLNGQTLKWVDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536
   |||||
Db 481 LLSKSVQLNGQTLKWVDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536
   |||||

RESULT 3
ADT78176
ID ADT78176 standard; protein; 536 AA.
XX
AC ADT78176;
XX
DT 13-JAN-2005 (first entry)
XX
DE Rat heparanase protein.
XX
KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cytostatic; antiinflammatory; vulnerrary; antiarteriosclerotic;
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; rat.
XX
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT Binding-site 150..155
FT Binding-site /note= "Putative heparin binding site"
FT Binding-site 264..270
FT Binding-site /note= "Putative heparin binding site"
FT Binding-site 419..426
FT Binding-site /note= "Putative heparin binding site"
XX
US2004213789-A1.
PN
PD 28-OCT-2004.
XX

```

```

PF 22-AUG-2003; 2003US-00645659.
XX
PR 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 01-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
XX
PA (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (FEIN/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (GELD/) GELDER J M V.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
Ayal-Herskovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
Friedmann Y;
WPI; 2004-774790/76.
XX
PT New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
PS Claim 5; SEQ ID NO 3; 68pp; English.
XX
CC The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents rat heparanase.
XX
SQ Sequence 536 AA;
Query Match      100.0%; Score 2800; DB 8; Length 536;
Best Local Similarity 100.0%; Pred. No. 8.6e-260;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRPLLLWLGRLALQTQGTAGTAPTKDVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
   |||||
Db 1 MLRPLLLWLGRLALQTQGTAGTAPTKDVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
   |||||

QY 61 DPRFLTLGSPRLALAGLSPAYLRFGGTKTDFLIIDPNKEPTSEERSYQSQNDNDIC 120
   |||||
Db 61 DPRFLTLGSPRLALAGLSPAYLRFGGTKTDFLIIDPNKEPTSEERSYQSQNDNDIC 120
   |||||

QY 121 GSERVSADVLRLKQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
   |||||
Db 121 GSERVSADVLRLKQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
   |||||

QY 181 LNALLRTPDLRWNSSNAQLLNTCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   |||||

```

Db 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240
Qy 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Qy 301 EDFLSSDVLDTFILSVQKILKVTKEMTGPKKWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db 301 EDFLSSDVLDTFILSVQKILKVTKEMTGPKKWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Qy 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNENFELPDYWLISLLFKLVGPKVLMRSRVKGP 420
Db 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNENFELPDYWLISLLFKLVGPKVLMRSRVKGP 420
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSDVVDKYLKPFSGSDG 480
Db 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSDVVDKYLKPFSGSDG 480
Qy 481 LLSKSVQLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536
Db 481 LLSKSVQLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536
RESULT 4
ID ADY27035
AC ADY27035 standard; protein; 536 AA.
AC ADY27035;
DT 05-MAY-2005 (first entry)
DE Rat heparanase protein.
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW protease; enzyme; enzyme purification.
OS Rattus norvegicus.
XX WO2005016227-A2.
XX 24-FEB-2005.
XX 12-AUG-2004; 2004WO-IL000744.
XX 14-AUG-2003; 2003US-0494800P.
XX 12-JAN-2004; 2004US-0535492P.
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX Van-Gelder JM, Miron D;
XX WPI; 2005-182203/19.
XX Regulating heparanase activity, useful for treating heparanase-associated
XX diseases (e.g. cancer, inflammation, cardiovascular diseases,
XX neurological diseases or viral diseases) comprises modulating heparanase
XX activation.
XX Disclosure; SEQ ID NO 7; 211pp; English.
XX The invention relates to a method of regulating heparanase activity in a
XX tissue or regulating a biological process depending at least in part on
XX heparanase activity comprising modulating heparanase activation. The
XX invention also relates to methods of treating a heparanase- or heparin
XX binding protein-associated disease or disorder in a subject, a
XX pharmaceutical composition for use in the treatment of a heparanase-
XX associated disease or disorder comprising a therapeutic amount of an
XX agent capable of modulating heparanase activation and a pharmaceutical
XX carrier or diluent, a method of identifying a protease activator of
XX heparanase, a protease substrate mimetic comprising a peptide

CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a rat heparanase protein used in the scope of the invention.
XX Sequence 536 AA;
Query Match 100.0%; Score 2800; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 8.6e-260;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLRPLLLLLWGRRLALTOGTTPAGTAPKDVVDLEFYTCKRLFQSVSPSLTITDASLAT 60
Db 1 MLRPLLLLLWGRRLALTOGTTPAGTAPKDVVDLEFYTCKRLFQSVSPSLTITDASLAT 60
Qy 61 DPRFLTEFLGSPRLRALARGLSPAYLRFEGGTTKDFLIFDPNKEPTSEERSYQSQDNNDIC 120
Db 61 DPRFLTEFLGSPRLRALARGLSPAYLRFEGGTTKDFLIFDPNKEPTSEERSYQSQDNNDIC 120
Qy 121 GSERVSADVLRLQMEWPFOELLLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
Db 121 GSERVSADVLRLQMEWPFOELLLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
Qy 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240
Db 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240
Qy 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Qy 301 EDFLSSDVLDTFILSVQKILKVTKEMTGPKKWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db 301 EDFLSSDVLDTFILSVQKILKVTKEMTGPKKWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Qy 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNENFELPDYWLISLLFKLVGPKVLMRSRVKGP 420
Db 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNENFELPDYWLISLLFKLVGPKVLMRSRVKGP 420
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSDVVDKYLKPFSGSDG 480
Db 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSDVVDKYLKPFSGSDG 480
Qy 481 LLSKSVQLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536
Db 481 LLSKSVQLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536
RESULT 5
AEA42425
ID AEA42425 standard; protein; 536 AA.
XX AEA42425;
AC AEA42425;
XX 28-JUL-2005 (first entry)
XX Rat heparanase epitope peptide SEQ ID NO:3.
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
XX antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
XX inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
XX angiogenesis disorder; cancer; tumor; metastasis.
OS Rattus norvegicus.
XX AU2004201462-A1.
XX 06-MAY-2004.
XX

PF 08-APR-2004; 2004AU-00201462.
XX
PR 08-APR-2004; 2004AU-00201462.
XX
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
PI Feinmstein Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinmstein E;
XX
XX WPI; 2005-173343/19.
XX
XX Novel isolated antibody capable of specifically binding to epitope of
PT heparanase protein, useful for preventing and treating heparanase-related
PT disorder such as inflammatory disorder, scars, autoimmune conditions or
PT angiogenesis.
XX
XX Claim 2; SEQ ID NO 3; 260pp; English.
XX
XX The invention relates to an isolated antibody or its portion (I) capable
CC of specifically binding to an epitope of a heparanase protein. Also
CC described: (1) a cell line (II) for producing a monoclonal antibody or
CC its portion, comprising a cell line for producing (I); (2) a
CC pharmaceutical composition comprising (I) and a carrier; and (3) an
CC affinity medium (III) for binding human heparanase polypeptides,
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
CC useful for treating a subject suffering from a pathological condition,
CC which involves administering (I) to the subject. (I) is useful for
CC preventing and treating heparanase-related disorder or condition chosen
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
CC cell proliferation, invasion of circulating tumor cells and metastatic
CC disease. (I) is useful for detecting the presence of heparanase
CC polypeptide in a sample. (I) is useful for detecting heparanase-related
CC disease or condition in a subject such as vertebrate, preferably mammal
CC e.g., human. The heparanase-related disorder or condition further
CC includes renal disease or disorder chosen from diabetic nephropathy,
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
CC and renal cell carcinoma. The present sequence represents rat heparanase,
CC which is used in the exemplification of the present invention.
XX
SQ Sequence 536 AA;

Query Match 100.0%; Score 2800; DB 9; Length 536;
Best Local Similarity 100.0%; Pred. No. 8.6e-260;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRPLLLMLWGRRLALTOGTAGTAPTKDVVDLEFYTKRLFQSVSPFSLITDASLAT 60
DB 1 MLRPLLLMLWGRRLALTOGTAGTAPTKDVVDLEFYTKRLFQSVSPFSLITDASLAT 60
QY 61 DRPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSDNDIC 120
DB 61 DRPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSDNDIC 120
QY 121 GSRVSADVLRLQMEWPFQELLRLROYQREFKNSTYSSSDVMLYSPAKCSRLDLIFG 180
DB 121 GSRVSADVLRLQMEWPFQELLRLROYQREFKNSTYSSSDVMLYSPAKCSRLDLIFG 180
QY 181 LNALRTPDLRWNSNAQLLLNYCSSKGYNISWELGNPNFSFWKKAQISIDGLQGEDFV 240
DB 181 LNALRTPDLRWNSNAQLLLNYCSSKGYNISWELGNPNFSFWKKAQISIDGLQGEDFV 240
QY 241 ELHKLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
DB 241 ELHKLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
QY 301 EDFLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGGAPILLSNTFAAGFWMLD 360
DB 301 EDFLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGGAPILLSNTFAAGFWMLD 360
QY 361 KLGLSAQLGIEVVRQVFFGAGNYHLVDENFEPLPDYWLSPKLVGPKVMSRVKGPD 420

DB 361 KLGLSAQLGIEVVRQVFFGAGNYHLVDENFEPLPDYWLSPKLVGPKVMSRVKGPD 420
QY 421 RSKLRVYLHCTNVVTHPRYREGDLTLVYVNLNHNVTYKHLKLPMPMFSRPVDKYLKPGSGDG 480
DB 421 RSKLRVYLHCTNVVTHPRYREGDLTLVYVNLNHNVTYKHLKLPMPMFSRPVDKYLKPGSGDG 480
QY 481 LLSKSVQLNGOTLKWVDSOTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKTAACI 536
DB 481 LLSKSVQLNGOTLKWVDSOTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKTAACI 536

RESULT 6
ADY27033
ID ADY27033 standard; protein; 535 AA.
XX
AC ADY27033;
XX
DT 05-MAY-2005 (first entry)
XX
DE Murine heparanase protein.
XX
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW protease; enzyme; enzyme purification.
XX
OS Mus musculus.
XX
PN WO2005016227-A2.
XX
PD 24-FEB-2005.
XX
PF 12-AUG-2004; 2004WO-IL000744.
XX
PR 14-AUG-2003; 2003US-0494800P.
PR 12-JAN-2004; 2004US-0535492P.
XX
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PI Van-Gelder JM, Miron D;
XX
DR WPI; 2005-182203/19.
XX
XX
PT Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.
XX
XX Disclosure; SEQ ID NO 5; 211pp; English.
XX
CC The invention relates to a method of regulating heparanase activity in a
CC tissue or regulating a biological process depending at least in part on
CC heparanase activity comprising modulating heparanase activation. The
CC invention also relates to methods of treating a heparanase- or heparin
CC binding protein-associated disease or disorder in a subject, a
CC pharmaceutical composition for use in the treatment of a heparanase-
CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a murine heparanase protein used in the scope of the
CC invention.
XX
SQ Sequence 535 AA;

Query Match 92.7%; Score 2594.5; DB 9; Length 535;
Best Local Similarity 92.7%; Pred. No. 5.1e-240;
Matches 497; Conservative 19; Mismatches 19; Indels 1; Gaps 1;

QY 1 MLRPLLLWGLRRLALTOGTAGTAPTDVVDLEFYTKRFLFQSVSPSFLSITIDASLAT 60
DB 1 MLR-LLLLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59

QY 61 DRPFLTLGSPRLRALARGLSPAYLRFGGTYKTDPLIPDPNKEPTSEBSRWQSDNDIC 120
DB 60 DRPFLTLGSPRLRALARGLSPAYLRFGGTYKTDPLIPDPNKEPTSEBSRWQSDNDIC 119

QY 121 GSRVSADVLRLQMEWPFQELLRLREYQREKFNSTYSSVDMYLFSAKCSGLDLIFG 180
DB 120 RSEPVSAVLRLQMEWPFQELLRLREYQREKFNSTYSSVDMYLFSAKCSGLDLIFG 179

QY 181 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV 240
DB 180 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV 239

QY 241 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
DB 240 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
DB 300 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359

QY 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLNGPKVMSRVKGP 420
DB 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLNGPKVMSRVKGP 419

QY 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTKHLKLPMPFMRSPVDKYLKPFSGDG 480
DB 420 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTKHLKVPFPLFRKPVDTYLLKPSGPG 479

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536
DB 480 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSALSVPFASFGFFVIRNAKIAACI 535

RESULT 7

AAB08851
ID AAB08851 standard; protein; 535 AA.

AC AAB08851;

XX 15-JAN-2001 (first entry)

XX A murine heparanase polypeptide.

XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
XX heparin-binding growth factor; cytokine; neurodegenerative plaque;
XX wound healing; infection; burn; angiogenesis; restenosis;
XX atherosclerosis; inflammation; neurodegenerative disease;
XX Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.

XX Mus sp.

XX W0200052178-A1.

XX 08-SEP-2000.

XX 14-FEB-2000; 2000W0-US003542.

XX 01-MAR-1999; 99US-00258892.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.

XX Pecker I, Vlodavsky I, Feinstein E;

XX

DR WPI: 2000-579289/54.
DR N-PSDB; AAA75081.

XX New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.

PS Claim 22; Page 144-145; 152pp; English.

XX The present sequence represents murine protein with heparanase catalytic
CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC particularly in treating tumour, inflammation or autoimmunity.
CC Particularly, the polynucleotide is useful in modulating the
CC bioavailability of heparin-binding growth factors, cellular responses to
CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
CC susceptibility to certain viral and some bacterial and protozoa
CC infections, or disintegration of neurodegenerative plaques. The
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
CC radiation burns), and in the treatment of angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC bacterial or protozoa infections

XX Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 3; Length 535;
Best Local Similarity 92.5%; Pred. No. 1.2e-239;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 1 MLRPLLLWGLRRLALTOGTAGTAPTDVVDLEFYTKRFLFQSVSPSFLSITIDASLAT 60

DB 1 MLR-LLLLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59

QY 61 DRPFLTLGSPRLRALARGLSPAYLRFGGTYKTDPLIPDPNKEPTSEBSRWQSDNDIC 120

DB 60 DRPFLTLGSPRLRALARGLSPAYLRFGGTYKTDPLIPDPNKEPTSEBSRWQSDNDIC 119

QY 121 GSRVSADVLRLQMEWPFQELLRLREYQREKFNSTYSSVDMYLFSAKCSGLDLIFG 180

DB 120 RSEPVSAVLRLQMEWPFQELLRLREYQREKFNSTYSSVDMYLFSAKCSGLDLIFG 179

QY 181 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV 240

DB 180 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV 239

QY 241 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300

DB 240 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360

DB 300 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359

QY 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLNGPKVMSRVKGP 420

DB 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLNGPKVMSRVKGP 419

QY 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTKHLKLPMPFMRSPVDKYLKPFSGDG 480

DB 420 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTKHLKVPFPLFRKPVDTYLLKPSGPG 479

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASFGFFVIRNAKIAACI 536

DB 480 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSALSVPFASFGFFVIRNAKIAACI 535

RESULT 8

ABB07811

ID ABB07811 standard; protein; 535 AA.

XX ABB07811;

XX

DT 03-JUL-2002 (first entry)
 XX Mouse heparanase sequence.
 DE
 XX Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;
 KW anti-protozoan; neuroprotective; heparin; mouse.
 XX
 XX Mus musculus.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..17
 FT Protein /note= "putative signal peptide"
 FT Protein 18..535
 FT Protein /note= "mature protein"
 XX
 PN US2002034810-A1.
 XX
 XX 21-MAR-2002.
 XX
 XX 16-AUG-2001; 2001US-00930218.
 XX
 XX 20-SEP-2000; 2000US-00666390.
 XX
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
 XX
 XX WPI; 2002-338926/37.
 XX
 XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
 PT to treat various heparin-related disorders and the signal peptide is
 PT useful in production of membrane-targeted or secreted recombinant
 PT proteins.
 XX
 XX Disclosure; Fig 1a; 39pp; English.
 XX
 XX The invention relates to an isolated avian and reptile nucleic acid,
 CC encoding a polypeptide with heparanase catalytic activity. The signal
 CC peptide of the nucleic acid can be used to express membrane-associated or
 CC secreted proteins in heterologous expression systems. The encoded
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
 CC invasion, and to intervene with pathologies associated with impaired
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoa and bacterial infections or
 CC disintegration of neurodegenerative plaques. The present sequence
 CC represents a mouse heparanase protein sequence used in similarity studies
 XX
 XX Sequence 535 AA;
 XX
 Query Match 92.5%; Score 2590.5; DB 5; Length 535;
 Best Local Similarity 92.5%; Pred. No. 1.2e-239;
 Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;
 QY 1 MLRPLLLWLRALRATQGTAGTAPTKVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
 DB 1 MLR-LLLLWLGALQAQAGAPAGTPTDDVVDLEFYTKRLRPSVSPSFLSITIDASLAT 59
 QY 61 DPRFLTLGSPRLRALARGISPAVLRGGTGTDFLI FDPNKEPTSEERSYQSQNDNIC 120
 DB 60 DPRFLTLGSPRLRALARGISPAVLRGGTGTDFLI FDPDKEPTSEERSYQSQVNHDC 119
 QY 121 GSERVSADVLRLKQWEPFDELLLRQYQREKNTSYRSVDMLYSFACSKRLDIFG 180
 DB 120 RSEPVSAVLRKUGVQVEFPQLLRQYQKEFNKNTSYRSVDMLYSFACSKGLDIFG 179
 QY 181 LNALRTPDLRWNSNAQLLNYCSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
 DB 180 LNALRTPDLRWNSNAQLLNYCSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
 QY 241 ELHKLLQKSAFQNAKLYGPDIGQPRGKTVKLLRSLKAGGEVIDSLTWHHYLNGRVATK 300
 DB 240 ELHKLLQKSAFQNAKLYGPDIGQPRGKTVKLLRSLKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLGTSAYGGAPLISNTFAAGFMWLD 360
 DB 300 EDFLSSDALDTFILSVQKILKVTKEITPGKKVWLGTSAYGGAPLISNTFAAGFMWLD 359
 QY 361 KLGLSAQGLGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFLKLVGPKVLMRSRVKGP 420
 DB 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFLKLVGPRVLLSRVKGPD 419
 QY 421 RSKLRVYLHCTNVYHPRYREGDLTLYVLNLHNVTKHLKLPMPFSPVDKYLKPFSGDG 480
 DB 420 RSKLRVYLHCTNVYHPRYQEGDLTLYVLNLHNVTKHLKVPPELFRKPDVTYLLKPSGPD 479
 QY 481 LLSKSVQLNGOTLKMVDROTLPALTEKPLPAGSSLSVPAPSFYGFVIRNAKIAACI 536
 DB 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSYGFVIRNAKIAACI 535

RESULT 9
 ADG88834
 ID ADG88834 standard; protein; 535 AA.
 XX
 AC ADG88834;
 XX
 DT 11-MAR-2004 (first entry)
 XX
 DE Mouse hpa protein.
 XX
 KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
 KW necrosis; pressure wound; diabetic ulcer; angiogenesis; mouse; therapy.
 XX
 OS Mus musculus.
 XX
 PN US2003161823-A1.
 XX
 XX 28-AUG-2003.
 XX
 XX 14-JAN-2003; 2003US-00341582.
 XX
 XX 31-AUG-1998; 98WO-US017954.
 PR 01-MAR-1999; 99US-00258892.
 PR 06-FEB-2001; 2001US-00776874.
 PR 05-SEP-2001; 2001WO-IL000830.
 PR 19-NOV-2001; 2001US-00988113.
 XX
 PA (ILAN/) ILAN N.
 PA (VLOD/) VLODAVSKY I.
 PA (YACO/) YACOBY-ZEEVI O.
 PA (PECK/) PECKER I.
 PA (FEIN/) FEINSTEIN E.
 XX
 XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
 PI
 XX
 XX WPI; 2003-897910/82.
 DR N-FSDB; ADG88833, ADG88835.
 XX
 XX Composition for treating a wound comprising recombinant heparanase is
 PT useful to induce or accelerate wound healing and induce or accelerate
 PT angiogenesis.
 XX
 XX Claim 2; SEQ ID NO 44; 143pp; English.
 PS
 XX
 CC The present invention relates to methods and compositions for inducing
 CC and/or accelerating wound healing via the catalytic activity of
 CC heparanase. The invention is used to induce or accelerate a healing
 CC process, particularly of an ulcer, burn, laceration, surgical incision,
 CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate
 CC angiogenesis. The present sequence is mouse hpa protein.
 XX
 XX Sequence 535 AA;
 QY
 Query Match 92.5%; Score 2590.5; DB 7; Length 535;
 Best Local Similarity 92.5%; Pred. No. 1.2e-239;

Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	
QY	1 MLRPLLLWGLRALTOGTAGTAPTKOVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
Db	1 MLR-LLLLLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
QY	61 DRPFLTLGSPRLRALARGLSPAYLRFGGTTKDFLIDPNKPTSEERSYQSQDNNDIC 120
Db	60 DRPFLTLGSPRLRALARGLSPAYLRFGGTTKDFLIDPDKEPTSEERSYKWSQVNHIC 119
QY	121 GSERVSADVLRLQMEWPFQELLRLREYQYQEFKNSTYSRSSVDMLYSFACSKRLDLIFG 180
Db	120 RSEPVSAAVLRKLQVWPFQELLRLREYQYQEFKNSTYSRSSVDMLYSFACSKGLDLIFG 179
QY	181 LNALLTPTDLRWNSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
Db	180 LNALLTPTDLRWNSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
QY	241 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db	240 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299
QY	301 EDPLSSDALDTFLLSVQKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db	300 EDPLSSDALDTFLLSVQKILKVTKETTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
QY	361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPDPYWLKLLPKLVGPKVLMRSVKGPD 420
Db	360 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDENFEPDPYWLKLLPKLVGPRVLSRVKGPD 419
QY	421 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTGHLKLPMPFSPRPVDKYLLKPFSGDG 480
Db	420 RSKLRVYLHCTNVYHPRYQEGDLTLVYLNHNVTGHLKVPDPPLFRKPDVTYLLKPSGPDG 479
QY	481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI 536
Db	480 LLSKSVOLNGQILKMWDEQTLPALTEKPLPAGSALSLPAFSYGFFVIRNAKIAACI 535
RESULT 10	
ADLI6413	
ID	ADLI6413 standard; protein; 535 AA.
XX	
AC	ADLI6413;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Mouse heparanase protein.
XX	
KW	Mouse; heparanase; enzyme; heparanase-dependent cancer; cancer;
KW	autoimmune reaction; inflammation.
XX	
OS	Mus musculus.
XX	
PN	US2003236215-A1.
XX	
PD	25-DEC-2003.
XX	
PF	09-JUN-2003; 2003US-00456573.
XX	
PR	31-AUG-1998; 98WO-US017954.
PR	01-MAR-1999; 99US-00258892.
PR	08-NOV-1999; 99US-00435739.
XX	
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX	
PI	Pecker I, Vlodavsky I, Feinstein E;
XX	
XX	WPI; 2004-070610/07.
XX	
PT	New antisense oligonucleotide hybridizable with a polynucleotide encoding
pt	a polypeptide with heparanase activity, useful for treating diseases such

PT	as cancer and autoimmune disorders.
XX	
PS	Claim 3; SEQ ID NO 44; 108pp; English.
XX	
CC	The invention relates to an antisense oligonucleotide (ASO) comprising a
CC	polynucleotide or a polynucleotide analogue of at least 10 bases being
CC	hybridisable in vivo , under physiological conditions, with a portion of
CC	a polynucleotide strand encoding a polypeptide having heparanase
CC	catalytic activity. Also included are a method of in vivo downregulating
CC	heparanase activity (comprising administering the ASO in vivo), a method
CC	of treating a subject suffering from a pathological condition
CC	(characterised by heparanase activity, comprising administering ASO to
CC	the subject), a pharmaceutical composition comprising the ASO and a
CC	carrier, an antisense nucleic acid construct (comprising a promoter
CC	sequence and a polynucleotide sequence directing the synthesis of an
CC	antisense RNA sequence of at least 10 bases being hybridisable in vivo ,
CC	under physiological conditions, with a polynucleotide strand encoding a
CC	polypeptide having heparanase catalytic activity), a method of in vivo
CC	downregulating heparanase activity (comprising administering in vivo the
CC	antisense nucleic acid construct), a pharmaceutical composition
CC	comprising the antisense nucleic acid construct and a carrier, and an
CC	antisense oligonucleotide comprising a polynucleotide or a polynucleotide
CC	analogue of at least 10 bases being hybridisable in vivo , under
CC	physiological conditions, with a portion of a polynucleotide strand being
CC	characterised by forming at least a portion of an untranslated region
CC	(UTR) for a polynucleotide strand encoding a polypeptide having
CC	heparanase catalytic activity. The methods and compositions of the
CC	present invention are useful for the prevention and/or treatment of
CC	diseases or conditions associated with aberrant heparanase activity, such
CC	as heparanase-dependent cancer, cancer, autoimmune reaction and
CC	inflammation. The gene for human heparanase is located on chromosome 4.
CC	The present sequence is the mouse heparanase protein.
XX	
SQ	Sequence 535 AA;

Query Match	92.5%; Score 2590.5; DB 8; Length 535;
Best Local Similarity	92.5%; Pred. No. 1.2e-239;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	

QY	1 MLRPLLLWGLRALTOGTAGTAPTKOVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
Db	1 MLR-LLLLLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
QY	61 DRPFLTLGSPRLRALARGLSPAYLRFGGTTKDFLIDPNKPTSEERSYQSQDNNDIC 120
Db	60 DRPFLTLGSPRLRALARGLSPAYLRFGGTTKDFLIDPDKEPTSEERSYKWSQVNHIC 119
QY	121 GSERVSADVLRLQMEWPFQELLRLREYQYQEFKNSTYSRSSVDMLYSFACSKRLDLIFG 180
Db	120 RSEPVSAAVLRKLQVWPFQELLRLREYQYQEFKNSTYSRSSVDMLYSFACSKGLDLIFG 179
QY	181 LNALLTPTDLRWNSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
Db	180 LNALLTPTDLRWNSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
QY	241 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db	240 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299
QY	301 EDPLSSDALDTFLLSVQKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db	300 EDPLSSDALDTFLLSVQKILKVTKETTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
QY	361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPDPYWLKLLPKLVGPKVLMRSVKGPD 420
Db	360 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDENFEPDPYWLKLLPKLVGPRVLSRVKGPD 419
QY	421 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTGHLKLPMPFSPRPVDKYLLKPFSGDG 480
Db	420 RSKLRVYLHCTNVYHPRYQEGDLTLVYLNHNVTGHLKVPDPPLFRKPDVTYLLKPSGPDG 479
QY	481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI 536

DB 480 LLSKSVQLNGQILKWVDEQTLPALTEXPLPGSALSLSLPAFSYGFFVIRNAKIAACI 535

RESULT 11

ADM48750

ID ADM48750 standard; protein; 535 AA.

AC ADM48750;

XX 03-JUN-2004 (first entry)

XX Mouse hpa protein.

DE Transgenic animal; heparanase; cancer; viral infection; restenosis; neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa; mouse.

KW Mus musculus.

OS US2003217375-A1.

XX 20-NOV-2003.

XX 24-FEB-2003; 2003US-00371218.

XX 31-AUG-1998; 98WO-US017954.

PR 01-MAR-1999; 99US-00258892.

PR 06-FEB-2001; 2001US-00776874.

PR 19-NOV-2001; 2001US-00988113.

XX (ZCHA// ZCHARIA E.

PA (VLOD// VLODAVSKY I.

PA (METZ// METZGER S.

PA (PECK// PECKER I.

PA (ILAN// ILAN N. SHAUL T.

PA (CHAJ// CHAJEK-SHAUL T.

PA (GOLD// GOLDSCHMIDT O.

XX Zcharia E, Vladavsky I, Metzger S, Pecker I, Ilan N;

PI Chajek-Shaul T, Goldshmidt O;

XX WPI; 2004-021918/02.

DR N-PSDB; ADM48749, ADM48751.

XX New transgenic non-human animal expressing heparinase, useful as models for human disease, such as cancers, viral infection, neurodegenerative diseases, restenosis, atherosclerosis and pulmonary disorders.

XX Example 12; SEQ ID NO 44; 106pp; English.

XX The present invention relates to a transgenic non-human animal whose genome comprises an exogenous polynucleotide sequence, including a promoter active in tissues of the non-human, a region encoding a human heparanase, where the promoter and the region encoding human heparanase are operably linked in the exogenous polynucleotide such that human heparanase is expressed in at least a portion of the cells of the non-human animal. The methods and compositions of the present invention are useful for the production of transgenic animals expressing heparanase, to be used as models for human diseases such as cancers, viral infection, restenosis, neurodegenerative diseases, atherosclerosis and pulmonary disorders. The present sequence is mouse hpa protein used in the exemplification of the invention.

XX Sequence 535 AA;

XX Query Match 92.5%; Score 2590.5; DB 8; Length 535;

XX Best Local Similarity 92.5%; Pred. No. 1.2e-239;

XX Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

Qy 1 MLRPLLWLGRLALTOGTPAGTATKDVVDLEFYTKRLFQSVSPFSITIDASLAT 60

DB 1 MLR-LLLLWGLGALAQAGACTPTDDVDLEFYTKRPLRSVSPFSITIDASLAT 59

PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VGEL/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX
DR WPI; 2004-625084/60.
XX
PT Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX
PS Claim 2; SEQ ID NO 2; 59pp; English.
XX
CC The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an epitope or its portion
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease), for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC non-solid cancer. The solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is mouse heparanase.
XX
SQ Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 8; Length 535;
Best Local Similarity 92.5%; Pred. No. 1.2e-239;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

Db 180 LNALLRTPDLRWNSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
Qy 241 ELHKLLQKSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRVATK 300
Db 240 ELHKLLQKSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
Qy 301 EDFLSSDVLDTFFILSVQKILKVTKEMTGPKKKVWLGSETSSAYGGGAPLLSNTFAAGFMWLD 360
Db 300 EDFLSSDALDTFFILSVQKILKVTKEITPGKKVWLGSETSSAYGGGAPLLSNTFAAGFMWLD 359
Qy 361 KLGLSAQLGIEVVMQVFFGAGNVHLVDENPEPLPDYWLKLLFKLVGPKVLMRSVKGPD 420
Db 360 KLGLSAQMGIEVVMQVFFGAGNVHLVDENPEPLPDYWLKLLFKLVGPRVLLSRVKGPD 419
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHLNVTKHLKLPMPFSRPVDKYLKPFPGSDG 480
Db 420 RSKLRVYLHCTNVYHPRYQEGDLTLVYLNHLNVTKHLKVPPLPRKPVDTYLLKPSGPDG 479
Qy 481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536
Db 480 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSALSILPAFSYGFVIRNAKIAACI 535
RESULT 13
ADT78175
XX ID ADT78175 standard; protein; 535 AA.
XX AC ADT78175;
XX DT 13-JAN-2005 (first entry)
XX DE Mouse heparanase protein.
XX KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
XX cell proliferation; cancerous condition; tumour cell invasion;
XX metastatic disease; heparanase-related disorder; inflammatory disorder;
XX wound; scar; vasculopathy; autoimmune condition; renal disease;
XX cytostatic; antiinflammatory; vulneryary; antiarteriosclerotic;
XX vasotropic; immunosuppressive; nephrotropic; antidiabetic; mouse.
XX OS Mus musculus.
XX FH Key Location/Qualifiers
FT Binding-site 149..154 /note= "Putative heparin binding site"
FT Binding-site 263..269 /note= "Putative heparin binding site"
FT Binding-site 419..425 /note= "Putative heparin binding site"
XX US2004213789-A1.
XX PD 28-OCT-2004.
XX PF 22-AUG-2003; 2003US-00645659.
XX PR 02-SEP-1997; 97US-00922170.
XX PR 01-MAY-1998; 98US-00071739.
XX PR 04-NOV-1998; 98US-00186200.
XX PR 19-FEB-2003; 2003US-00368044.
XX PA (YACO/) YACOBY-ZEEVI O.
XX PA (PERE/) PERETZ T.
XX PA (MIRO/) MIRON D.
XX PA (SHLO/) SHLOMI Y.
XX PA (PECK/) PECKER I.
XX PA (AYAL/) AYAL-HERSHKOVITZ M.
XX PA (FEIN/) FEINSTEIN E.
XX PA (VGEL/) VAN GELDER J M V.
XX PA (VLOD/) VLODAVSKY I.
XX PA (FRIE/) FRIEDMANN Y.
XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX
XX
DR WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
XX
PS Claim 5; SEQ ID NO 2; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents mouse heparanase.

XX Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 8; Length 535;
Best Local Similarity 92.5%; Pred. No. 1.2e-239;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;
QY 1 MLRPLLLWLGRLALTOGTPAGTAPKDVLDLEFYTKRLFQSVSPSFLITDASLAT 60
DB 1 MLR-LLLLLWLGALQAQAGAPAGTAPTDVLDLEFYTKRPLRSVPSFLITDASLAT 59
QY 61 DPRFLPLGSPRLALARGLSPAYLRGGTKTDPLIEDPNKEPTSEBSRWQSDNDIC 120
DB 60 DPRFLPLGSPRLALARGLSPAYLRGGTKTDPLIEDPNKEPTSEBSRWKSNQVNDIC 119
QY 121 GSERVSADVLKQMEWPFQELLILLREQYQREFKNSTYSSVDMLYSFAKCSRLDLIFG 180
DB 120 RSEPVSAVLRKQVWEPFQELLILLREQYQREFKNSTYSSVDMLYSFAKCSGLDLIFG 179
QY 181 LNALLRTPDLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
DB 180 LNALLRTPDLRWNSNAQLLDDYCSSKGYNISWELGNEPNSFWKKAHLLDGLQGEDFV 239
QY 241 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVDLSLTWHYHLYNGRVATK 300
DB 240 ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVDLSLTWHYHLYNGRIATK 299
QY 301 EDFLSSDVLDTFLSVQKILKVTKEMTPGKVMWLGETSAYGGCAPLLSNTFAAGFWMLD 360
DB 300 EDFLSSDVLDTFLSVQKILKVTKEMTPGKVMWLGETSAYGGCAPLLSNTFAAGFWMLD 359
QY 361 KLGLSAQLGHEVVMRQVFFGAGNHLVDENFEPPLDYWLKLLFKLVGPKVLMRSVRKGPD 420
DB 360 KLGLSAQMGHEVVMRQVFFGAGNHLVDENFEPPLDYWLKLLFKLVGPKVLMRSVRKGPD 419
QY 421 RSKLRVYLHCTNVYHPYRSGDLTLYVLNLHNVTKHLKLPMPFMRPVDKYLLKPFSGSDG 480
DB 420 RSKLRVYLHCTNVYHPYRSGDLTLYVLNLHNVTKHLKLPMPFMRPVDKYLLKPFSGPDG 479

QY 481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASVGFVIRNAKTAACI 536
DB 480 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSALSPLFASVGFVIRNAKTAACI 535

RESULT 14
AEA42424

ID AEA42424 standard; protein; 535 AA.

XX AC AEA42424;

XX DT 28-JUL-2005 (first entry)

XX DE Mouse heparanase epitope peptide SEQ ID NO:2.

XX antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
KW angiogenesis disorder; cancer; tumor; metastasis.

XX OS Mus musculus.

XX PN AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

PI Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinstein E;

XX WPI; 2005-173343/19.

PT Novel isolated antibody capable of specifically binding to epitope of
PT heparanase protein, useful for preventing and treating heparanase-related
PT disorder such as inflammatory disorder, scars, autoimmune conditions or
XX angiogenesis.

PS Claim 2; SEQ ID NO 2; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable
CC of specifically binding to an epitope of a heparanase protein. Also
CC described: (1) a cell line (II) for producing a monoclonal antibody or
CC its portion, comprising a cell line for producing (I); (2) a
CC pharmaceutical composition comprising (I) and a carrier; and (3) an
CC affinity medium (III) for binding human heparanase polypeptides,
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (1)
CC useful for treating a subject suffering from a pathological condition,
CC which involves administering (I) to the subject. (I) is useful for
CC preventing and treating heparanase-related disorder or condition chosen
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
CC cell proliferation, invasion of circulating tumor cells and metastatic
CC disease. (I) is useful for detecting the presence of heparanase
CC polypeptide in a sample. (I) is useful for detecting heparanase-related
CC disease or condition in a subject such as vertebrate, preferably mammal
CC e.g., human. The heparanase-related disorder or condition further
CC includes renal disease or disorder chosen from diabetic nephropathy,
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
CC and renal cell carcinoma. The present sequence represents mouse
CC heparanase, which is used in the exemplification of the present
XX invention.

XX SQ Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 9; Length 535;

Best Local Similarity 92.5%; Pred. No. 1.2e-239;		Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	
QY	1	MLRPLLLMLWGRRLALTOCTPAGTAPKDVVDLEFYTKELFQSVSPSFLSITIDASLAT	60
DB	1	MLR-LLLLMLWGLGALAQAGPAGTAPDDVDLEFYTKPLRSVSPSFLSITIDASLAT	59
QY	61	DRPFLTFLGSPRLRALARGLSPAYLRPGGTTKDFLIPDPNKEPTSEERSYQSQDNDIC	120
DB	60	DRPFLTFLGSPRLRALARGLSPAYLRPGGTTKDFLIPDPNKEPTSEERSYQSVNHIC	119
QY	121	GSESVADVLRKLQMEWPFQELLRLREOYQREPKNSTYSSSDVMLYSPAKCSRLDIFG	180
DB	120	RSEPVSAAVLRKLQMEWPFQELLRLREOYQREPKNSTYSSSDVMLYSPAKCSGLDIFG	179
QY	181	LNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV	240
DB	180	LNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	239
QY	241	ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK	300
DB	240	ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK	299
QY	301	EDPLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFWMLD	360
DB	300	EDPLSSDALDTFILSVQKILKVTKETTPGKKVWLGETSSAYGGAPLLSNTFAAGFWMLD	359
QY	361	KLGLSAQLGIEVVMROVFFGAGNYHLVDENFEPDPYWLSSLFKKLGPVKLMSRVKGP	420
DB	360	KLGLSAQMGIEVVMROVFFGAGNYHLVDENFEPDPYWLSSLFKKLGPVLLSRVKGPD	419
QY	421	RSKRLVYLHCTNYYHPRYREGDLTYVNLNHNVTKHLKLPMPMFSRPVDKYLKPFSGDG	480
DB	420	RSKRLVYLHCTNYYHPRYQEGDLTYVNLNHNVTKHLKVPDPLFRKPDVTYLLKPSGPDG	479
QY	481	LLSKSVOLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536
DB	480	LLSKSVOLNGQILKMVDEQTLPALTEKPLPAGSALSVPAFSYGFFVIRNAKIAACI	535
RESULT 15			
ADY27034			
ID	ADY27034 standard; protein; 545 AA.		
AC	ADY27034;		
XX	05-MAY-2005 (first entry)		
DT	Bovine heparanase protein.		
DE	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;		
XX	neurological disease; viral infection; infection; cytostatic;		
KW	antiflammatory; cardiovascular-gen.; neuroprotective; virucide;		
KW	protease; enzyme; enzyme purification.		
XX	Bos taurus.		
OS	WO2005016227-A2.		
XX	24-FEB-2005.		
XX	12-AUG-2004; 2004WO-IL000744.		
PP	14-AUG-2003; 2003US-0494800P.		
XX	12-JAN-2004; 2004US-0535492P.		
PR	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.		
XX	Van-Gelder JM, Miron D;		
PI	WPI; 2005-182203/19.		
XX	Regulating heparanase activity, useful for treating heparanase-associated		
DR			
XX			
PT			

PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,	
PT	neurological diseases or viral diseases) comprises modulating heparanase	
PT	activation.	
XX	Disclosure; SEQ ID NO 6; 211pp; English.	
XX	The invention relates to a method of regulating heparanase activity in a	
CC	tissue or regulating a biological process depending at least in part on	
CC	heparanase activity comprising modulating heparanase activation. The	
CC	invention also relates to methods of treating a heparanase- or heparin	
CC	binding protein-associated disease or disorder in a subject, a	
CC	pharmaceutical composition for use in the treatment of a heparanase-	
CC	associated disease or disorder comprising a therapeutic amount of an	
CC	agent capable of modulating heparanase activation and a pharmaceutical	
CC	carrier or diluent, a method of identifying a protease activator of	
CC	heparanase, a protease substrate mimetic comprising a peptide	
CC	representing a subset or all substrate residues or cleavage sites of	
CC	human heparanase or an equivalent non-human heparanase, a method of	
CC	producing active heparanase and a method of modulating an adhesion	
CC	activity of heparanase. The composition and methods are useful for	
CC	modulating heparanase activation and for treating heparanase-associated	
CC	diseases or disorders such as cancer, inflammation, cardiovascular	
CC	diseases, neurological diseases or viral infections. This sequence	
CC	represents a bovine heparanase protein used in the scope of the	
CC	invention.	
XX	Sequence 545 AA;	
QY	Query Match 76.8%; Score 2151; DB 9; Length 545;	
DB	Best Local Similarity 76.4%; Pred. No. 2.5e-197;	
DB	Matches 410; Conservative 50; Mismatches 75; Indels 2; Gaps 1;	
QY	2 LRPL-LLLLMLWGRRLALTOCTPAGTAPKDVVDLEFYTKELFQSVSPSFLSITIDASLA	59
DB	9 LRPLLLLLPLGLPGCSGTPAAAAPADDAELEFFTERPLHLVSPAFSLFTIDANLA	68
QY	60 TDRPFLTFLGSPRLRALARGLSPAYLRPGGTTKDFLIPDNKEPTSEERSYQSQDNDI	119
DB	69 TDRPFLTFLGSSKLRTLARGLAPAYLRFGNGKGDFLIFDPKKEPAFEERSYWLQSQNDI	128
QY	120 CGSERVADVLRKLQMEWPFQELLRLREOYQREPKNSTYSSSDVMLYSPAKCSRLDIF	179
DB	129 CKSGSPSDVEEKLRLLEWPFQELLRLREOYQREPKNSTYSSSDVMLYTFASCGLNLIF	188
QY	180 GLNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDF	239
DB	189 GVNALLRTPDWHDDSSNAQLLNTYSSKNTNISWELGNEPNSFORKACIFINGQLGDF	248
QY	240 VELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVAT	299
DB	249 IEFRLKLGSAFNAKLYGPDIGQPRENTVMKLSFLKAGGEVIDSVTWHHYVNGRIAT	308
QY	300 KEDFLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFWML	359
DB	309 KEDFLNPDIILDTFISVQKTLRIVEKIRPLKKVWLGETSSAFGGGAPPLSNTFAAGFWML	368
QY	360 DKGLSAQLGIEVVMROVFFGAGNYHLVDENFEPDPYWLSSLFKKLGPVKLMSRVKGP	419
DB	369 DKGLSARMGIEVVMROVLFAGNYHLVDGNFEPDPYWLSSLFKKLGVNKLMSVMSVKG	428
QY	420 DRSKRLVYLHCTNYYHPRYREGDLTYVNLNHNVTKHLKLPMPMFSRPVDKYLKPFSGD	479
DB	429 DRSKRVYLHCTNWKHPRYEGDLTYALNHNVTKHLPHLPHLHNKQVYKYLKPSGTD	488
QY	480 GLLSKSVOLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536
DB	489 GLLSKSVOLNGQILKMVDEQTLPALTEKPLHPSGSLGMPPEFSYGFFVIRNAKVAACI	545
Search completed: June 5, 2006, 12:09:42		
Job time : 107.86 secs		

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.8454 Seconds
(without alignments)
2889.939 Million cell updates/sec

Title: US-10-645-659A-3
Perfect score: 2800
Sequence: 1 MLRPLLLMLWGLRALTQG.....VPAFSYGFVIRNAKIAACI 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: Piri:*
2: Piri2:*
3: Piri3:*
4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	904	32.3	480	2 JC7506	heparanase protein
2	410	14.6	521	2 T45608	hypothetical prote
3	162.5	5.8	190	2 T01953	hypothetical prote
4	109	3.9	575	2 T45668	hypothetical prote
5	109	3.9	1331	2 A48954	mannan endo-1,4-be
6	108.5	3.9	4436	2 E71086	hypothetical prote
7	108	3.9	725	2 S01042	aerobactin recepto
8	107	3.8	837	1 A31842	endo-1,4-beta-xyla
9	106.5	3.8	1573	2 T50113	3-dehydroquinatase
10	106	3.8	599	2 S75363	hypothetical prote
11	106	3.8	2165	1 RRNZ42	genome polyprotein
12	105.5	3.8	575	2 T12094	beta-fructofuranos
13	104.5	3.7	1482	2 S13495	pregnancy zone pro
14	104	3.7	511	2 S61166	probable membrane
15	104	3.7	732	2 T44483	receptor-like prot
16	104	3.7	796	2 D97065	transketolase [imp
17	103.5	3.7	844	2 T52396	formin-binding pro
18	103.5	3.7	914	2 B96592	hypothetical prote
19	103.5	3.7	1462	2 T06819	DNA topoisomerase
20	103	3.7	361	2 A43510	basic membrane pro
21	102.5	3.7	587	2 S6231	beta-fructofuranos
22	102	3.6	879	2 F86875	probable fibribral
23	102	3.6	879	2 E91031	probable outer mem
24	102	3.6	1392	1 YGBYAD	L-aminoadipate-sem
25	101.5	3.6	412	2 G81581	tyrosyl-tRNA synth
26	101	3.6	581	2 E90449	conserved hypothet
27	101	3.6	596	2 T04506	hypothetical prote
28	101	3.6	709	2 A35364	carcinoembryonic a
29	100.5	3.6	495	2 C89778	hypothetical prote

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

C:Comment: This protein, an intracellular membrane-bound enzyme, has biological and therape

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match	32.3%	Score	904;	DB	2;	Length	480;
Best Local Similarity	37.8%	Pred. No.	6.1e-62;				
Matches	202;	Conservative	74;	Mismatches	150;	Indels	108;
Gaps	8;						
Qy	12	GRLRALTQGT	PAGTAPT	KDVVDLEFYTKRL	FQSVSPFLSITIDASLATDPRFLTLGSP	71	
Db	42	GDRPLPVDRAAG	-LKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHD	-GWLDFLSSK	99		
Qy	72	RLRALARGLS	PAYLRF	GGTKTDFLIFD	PNKEPTSEERSYWSQDNNDICGSRVSADVL	131	
Db	100	RLVTLARGLS	PAFLRFGGKRTDFLQF	-----	-----	125	
Qy	132	KLQWEPQELL	LLREQYQREFKSTYSRSSVDMLYS	PAKCSRLDLIFGLNALLRTPDLR	191		
Db	126	-----	-----	QNLRNPAKRGSGPG	-----	PD--	141
Qy	192	WNSSNAQLLL	NYCSSKGYNISWELCNEPNSFWKKAQISIDGLQ	LGEDFVELHKLQK-SA	250		
Db	142	-----	-----	YLLKNY-----	EDENNYVTMHGRAVNSQLGKDYIQLKSLQPIR	183	
Qy	251	FQNAKLYGPD	TGQPRGKTVKLLRSL	FAGGVIDSLTWHYHYNGRVATKE	DFLSSDVLJD	310	
Db	184	YSRASLYG	PNIGRPRKNVIALD	GGPMKVGSTDAVTWQHCVIDGRVVKVMDFL	KTRLDD	243	
Qy	311	TFILSVQKIL	KVTKEMTPGKVKWLGETSSA	GGGAPLISNTFAAGFMWLDKGLS	QLGI	370	
Db	244	TLSDQIRKIQ	IKVAVNTYTPGKKI	WLEGVVTTSGAGTNNLSDSYAAGFLNLT	IGMLANQGI	303	
Qy	371	EVVMRQVFG	AGNVLVDENFEPLDYWL	LSLLFKLVGPKVLMRSVK	GGPD-----	R	421
Db	304	DVIRHSPFD	HGYNHLVDQNFENPLPDYWL	SLDKKILGPKVLAVH	VAGLQRPGRVIR	363	

```

Qy 422 SKRVYLHCNTNYYHPRVREGDLTLVLNLHNVTXHLKLPMPMFSPRPVSKYLLKPFSGDGL 481
Db 364 DKLRVIAHCTNHHNHNHNVGRSGITLFIINLHRSRKKIKLAGTLTRDKLVHGYLLQPYGQEG 423
Qy 482 LSKSVQLNGQTLKMWDEOTLPALTEKPIIPAGSSLSVPAFSYGFFVIRNAKIIAAC 535
Db 424 KSKSVQLNGQPLVMVMDGTLPDLKPRPLRAGRTLIVPVMTGMFFVVKNNVALAC 477

RESULT 2
T45608
Hypotheical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T45608
R:Revan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckae
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UP100000AA497C; EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A:Note: F13G24.30

```

Query Match	14.6%;	Score 410;	DB 2;	Length 521;
Best Local Similarity	29.4%;	Pred. No. 1.1e-23;		
Matches 155;	Conservative	73;	Mismatches 179;	Indels 120; Gaps 25;
Qy	68	LGSPRLRALARGLSPAVLREGGKTDFLIFDPNKEPTSEERSYQWSDNNDICGSERVSA	127	
Db	55	LTRPLTTKAKPKPURIKIRIGGSQQQVIVDVGNLKT-----CR-----	94	
Qy	128	DVLRKLQMEWPFQELLLRLREQYQREKNS---TVSRSSV-----DMLYSPAKCSRLLDIF	179	
Db	95	-----PFQKM-----NSGLGFGSKGLHMKRWDELNSFLTATGAVVTF	132	
Qy	180	GLNALLRTPDLR-----WNSSNAQLLLNYCSGKGYNI-SWELGNPNPFWKCAQIAISD	231	
Db	133	GLNALRGHRKLRGKAWGCAWDHINTQDFLNTVTSKGYVIDSWEPFEGNELSGSGVGASVSAE	192	
Qy	232	GLQIGEDFVELHKLLOKSAFQNAKLYCPDICQPRG-----KTVKLLRSFLKAGEVIDLSL	286	
Db	193	--LYGKDILVILDKDINK-VYKNSLWHLKPIILVAPGGFYEQQWYTKLLBI---SGPSVVDVV	246	
Qy	287	TWHYYLNGRVATKEDFLSDVDL-TRFLSYQKILK----VTKEMTPGKVKWLCETSSAY	341	
Db	247	T-HHYINLG--SGNDPALVKIMDPISVLSQVSKTFKDVNQTIQEHGPAWSPWGESGGAY	303	
Qy	342	GGGAPLLSNTFAAGFMWLDKLGLSAQLGIEVYVRQVFFGAGNYHLVDE-NPEPLPDYWLS	400	
Db	304	NSGGRHVSDTFIDSFWYLDQLGMSARHTKVCYCRQTLVG-GFYGLLEKGTFPVNPDPYVA	362	
Qy	401	LLFPKLVGPVKVLMRSVKGPDRSKRLRVYLHCTNVYHPRYREGDULTYVLNLHN-----V	453	
Db	363	LLWHRLMGKGLVALQVTDGP--POLRVYAHCS-----KGRAG-VTLILINLSNQSDFTVSV	414	
Qy	454	TKHLKLPMPFSPVDKYLL---KPF-----GSDGLL-----SKSVQL	488	
Db	415	SGNINVVLNAESRK-KSILDLTKRPFSSWIGSKASDGYLANREEYHLTPENGVLRSKTMVL	473	
Qy	489	NGOTLKMDVQDTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKTAAC	535	
Db	474	NGKSLKPTAGDIPSLPEPLVRSVNSPNLVLPLSMFSFVLNPFNDASAC	520	

RESULT 3
T01953

hypothetical protein T2L5.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C:Accession: T01953
R:Geisel, C.; Smith, A.; Le, T.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana T2L5.
A:Reference number: Z14470
A:Accession: T01953
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-190 <GEI>
A:Cross-references: UNIPROT:O82604; UNIPARC:UPI000000A9F7D; EMBL:AF096371; NID:g3695386;
A:Experimental source: cultivar Columbia

Query Match	5.8%;	Score 162.5;	DB 2;	Length 190;	
Best Local Similarity	28.6%;	Pred. No. 3.3e-05;			
Matches	54;	Conservative 34;	Mismatches 62;	Indels 39; Gaps 10	
Qy	375	ROVFFGAGNHLVD-ENFEP	LPDYLSLFPKLVGPKVLM	SRVGPDRSLRVYLHCTNV 433	
Db	12	QSLIG-GNGLNTNTFTNP	DYYSALIRQLMGRKALFT	FTFSG--TKKIRSYTHCA-- 66	
Qy	434	YHPRYREGDUTLVYLN	LHNV-----TKHLK----	LPPMFSRP----- 467	
Db	67	---RQSKG-ITVLIMN	LDNTTTVVAKVELNNSF	SLRHTKHKMSYKRASSQLFGGPN	GVIQ 122
Qy	468	VDKYLKPFSGDG-LLSKS	VLNQNLTKWVDEQTLPAL	TEKPLPAGSSLSVPAPSYGFFV	526
Db	123	REEVHLT--AKDGNLH	SQTMLLNGNALQVNSMGD	LPPIEPHINSTEPIITAPISIVFH	180
Qy	527	IRNAKTAAC	535		
Db	181	MRNVVPAC	189		

RESULT 4
T45668
hypothetical protein F14P22.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T45668
R:D'Angelo, M.; Verzzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.
Submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23011
A:Accession: T45668
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-575 <DAN>
A:Cross-references: UNIPROT:Q9M2G8; UNIPARC:UPI000009B66E; EMBL:AL137082
A:Experimental source: cultivar Columbia; BAC clone F14P22
C:Genetics:
A:Map position: 3
A:Introns: 2/1; 126/2; 164/3; 186/3; 216/2; 245/3; 325/3; 359/3
A:Note: F14P22.70
C:Superfamily: Arabidopsis thaliana hypotheical protein F8J2.40

```

Query Match      3.9%; Score 109; DB 2; Length 575;
Best Local Similarity 22.9%; Pred. No. 2.4;
Matches 102; Conservative 57; Mismatches 140; Indels 146; Gaps 25

Qy 38 TKRLFGQVSPSFLSITIDASLATDP-----RFLTFLGSPRRALARGLSPAYLRFGG 89
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 3 TKTLRISi--SFNDIDSDSTITESPEARCTLNRSLSGKGTQSR-----ISP-----GD 50

Qy 90 TKTDFLI--FDPNKEPTSEERSYQWQSQNDIDCGSERVSADVLRLQWEPFOELLIRE 147
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 51 VENNFSIKPLTFKWKEDDRYKMRIRWK-----PVCNEEH-AKEFLALLGGDTYQAALKLOK 104
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

```

QY 148 QYREFKNTSYRSRSD-----MLYSFAKCSRLDLIFGLNALLRTPDLRWSSNA 197
Db 105 VY-RSFT---RRRLADCAVVEQRRWKVLDFAELKSSISFFIEKEQETAVSWRSART 160
QY 198 QLLNNYSSKGYNIS-----WELGNEPN-----SPWKAQIISIDGLQLGED 238
Db 161 R-----AAKVGKGLSKDEKARKLALQHWLEAIDPRHRYGHNLPYVYHAWLHCD----- 208
QY 239 FVELHKLLOKSAFONAKLYGPDICQ-----PRGK-----TVKLRSFLKAGEVIDS 285
Db 209 -----SKQPF-----FYWLDIGQKELNHERCPRSKLYQQSILKYLGPTEREAYEYIIE 256
QY 286 LTHWHYINGRVATKEDFLSSDVLDT-----FILSVQIKLVTKEMTPCKKWLWG 335
Db 257 -----DGKIMYKQ---SGVLDTKGPPDAKWIIFVLSVSKILYVGM-----KKGNF 300
QY 336 ETSAYGGGAPLLSNTFAAGFMWLDKLGSAQLGIEVVMRVFFGAGNYHLVDENFEPLP 395
Db 301 QHSFFLAGA-----TUSAGRIVD-----DGVLKAVWPHSGHYLPTEENFOA-- 343
QY 396 DYWLSLFFKLVGPKVLMRSVKGPD 420
Db 344 --FWSFLRENNVD---LANVKKNPD 363
RESULT 5
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
C:Species: Caldocellum saccharolyticum
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A:Reference number: A48954; MUID:93119139; PMID:1476429
A:Accession: A48954
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1331 <GIB>
A:Cross-references: UNIPROT:P22533; UNIPARC:UPI000012EB88; GB:L01257; NID:g144290; PIDN:
R:Luethi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A:Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A:Reference number: A43745; MUID:91247819; PMID:2039230
A:Accession: B43745
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337, 'PPRQHQHQ' <LUB>
A:Cross-references: UNIPARC:UPI000016EB78; EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PI
A:Note: the authors translated the codon CAC for residue 262 as Glu
A:Note: this sequence has been revised in reference A48954
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 3.9%; Score 109; DB 2; Length 1331;
Best Local Similarity 18.3%; Pred. No. 8.4;
Matches 104; Conservative 75; Mismatches 196; Indels 192; Gaps 23;
QY 21 TPAGTA-PTKDVVDLEFYTKRLFSQSPFSLSITIDASLATDPRFLTFLGSPRLAL--- 76
Db 758 TPTVTATPTPTPIPTVPTLPTISPSVSVVEITINTNAGRTQI-----SPIYIGANQD 811
QY 77 ARGLSPAYLRFGGTKTDFLIDPNKPTSEERSYVWQSQDNDICGS-----ERVADV 129
Db 812 IEGVHSARLGGNRLTGYNNWENN---FSNAGNDWYHSSDDYLCWSMGISGEDAKVPAV 868
QY 130 LRKLQWMPPEQLLLREQYQREFKNTSYRSRSDVLMYSFAKCS----- 173
Db 869 VSKF-----HEYSLKNAYSATVLQMGAGYVSKONYGTVSENETHAPSRWA 913
QY 174 ----RLDLIFGLNALLRTPDLRWSSNAQLLLNYCSSKGYNIS-----WELGNEPN 221

Db 914 EVFKCKDAPLSLN-----PDLNDFVYMDBEFINVLYNK-YGMASSTPGTGIKYILDNEPD- 966
QY 222 FWKKAQIISIDG-----LQGEDFVELHKLLOKSAFONAKLYG-----PDI 261
Db 967 LWASTHPRIHPNKVTCRELIEKSVELAKVI-KTLDPSAEVFGYASYGMYYSLODAPDW 1025
QY 262 GQPRGKTVKLLRSFLKA-----GGEVIDSLTWHHY----- 291
Db 1026 NOVGEHRWFISWLEQMKKASDSFGKELLDVLDLHWYPEARGGNIRVCFDGENDTSKEV 1085
QY 292 -----YLNGRVATKEDFSSDVLDTFILSVQIKLVTKEMTPCKKYM 333
Db 1086 VIARMQAPRTLWDPTYKTSVKGQITAGENSINQWFSVDYLPPIPNVKADIEKYVPTCKLA 1145
QY 334 LGETSSAYGGGAPLLSNTFAAGFMWLDKLGSAQLGIEVVMR-----QVF--- 378
Db 1146 ISEPD--YGG-----RNHISGGIALADVLFGKYGVNFAARWGDGSGSYAAAYNIYLN 1198
QY 379 -----FCAGNYHLVDENFEPLPDYWLSSLFKLVGPKVLMRSVKGPDRLVYLHCTN 432
Db 1199 DGKSKYGNNTVNSANTSDVENMPVY-----ASINGQDDSELHILINRN 1242
QY 433 VYHPRYREGDLTYV-LNLHNVTKHLK 458
Db 1243 Y-----DQKLQVKINITSTPKYTK 1261
RESULT 6
E71086
hypothetical protein PH0954 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: E71086
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine,
M.; Ohtoku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic al
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: E71086
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-4436 <KAW>
A:Cross-references: UNIPROT:O58659; UNIPARC:UPI00001100E3; GB:AP000004; NID:g3236131; PI
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0954
Query Match 3.9%; Score 108.5; DB 2; Length 4436;
Best Local Similarity 19.6%; Pred. No. 56;
Matches 101; Conservative 73; Mismatches 201; Indels 139; Gaps 22;
QY 77 ARGLSPAYLRFGGTKTDFLIDPNKPTSEERSYVW-QSQDNDICGSERVSADVLRLQ 134
Db 1018 SRDVTLYRLPSSGKHNLISVLDPYNR-----WIEENENNNLTFLSLFGKPKDLKVEG 1069
QY 135 MEW-----PFOELLRLREQYQREFKNT---YSRSSVDMLYSF 169
Db 1070 ITWAPYNTSGENVLFYIVKVLQOPFLKSTVR-----AEIWNCTRKIYSTNAYPRNWSF 1125
QY 170 AKCSRLDLIFGLNALLRTPDLRWSSNA-----QLLLNYCSS-----KGYNISWEIG 216
Db 1126 GKG-----ETKEFNWRWYNAKPGNLTKIVVDYVNSIPEGNESNNFSAPLG 1172
QY 217 NEPNFSFWKQAQISIDGLQGEDFVELHKLLOKSAFONAKLYGPDIGQPRGKTVKLLRSFL 276
Db 1173 NVGTPDFKLENLSVEDLAYGK-FVRIN-----ATVKNLGDsi 1208
QY 277 KAGEVIDSLTWHHYLNGRVATKEDFLSSDVLDTFILSVQIKLVTKEMTPCKKWLGE 336
Db 1209 YRPTVTLVFNVSGERYI-RTVYGIKENESKSVTLPHYVDRVGEV-RVKVEVDPRIVEGN 1266

```
QY 337 TSS-----AYGGAPLLSNTFAAGWMLDKGLLSAQLGIEVVMR-----QVFFGAGNYHLV 387
D 1267 ESNIIIRYYVESPEL---MLSGYEWLEEEVVRGYLAYKVNVTNTGQDVYFYQVMFV 1323
QY 388 DENFEPLDYLWLSLLFKKLGPVKLSRVKGPDRSKLRVYLHCTNVVHPYRREGDLTYV 447
D 1324 DG--EPKSSWINKLLHGTAERTLRWFPSSGGRKEVRIIVVDQD-YIPESNEDNNAI-- 1378
QY 448 LNLHNVTKHLKLPP-----PPMFS-----RPVDKYLKLPFGSGGLLS 483
D 1379 --VENVT--IVLPDIEVLSNIPSMHANSYFKVNATIKNSGGQDVKRIFYVSLYQDGKLL 1434
QY 484 KSVQLNGQTLKMVDEQTLPALTEKPLPAGSSLSV 517
D 1435 GSAPVYSLASGEVKEVT---LTIRPYGNSTFKV 1465

RESULT 7
S01042
A:Description: aerobactin receptor precursor iutA [validated] - Escherichia coli plasmid ColV-K30
N:Alternate names: cloacin receptor
C:Species: Escherichia coli
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: S01042
R:Krone, W.J.A.; Stegehuis, F.; Koningstein, G.; van Doorn, C.; Roosendaal, B.; de Graaf
FEMS Microbiol. Lett. 26, 153-161, 1985
A:Title: Characterization of the pColV-K30 encoded cloacin DF13/aerobactin outer membran
e sequence and primary structure.
A:Reference number: S01042
A:Accession: S01042
A:Molecule type: DNA
A:Residues: 1-725 <KRO>
A:Cross-references: UNIPROT:P14542; UNIPARC:UPI000017AA31; EMBL:X05874
C:Genetics:
A:Gene: iutA
A:Genome: plasmid ColV-K30
C:Function:
A:Description: functions as outer membrane receptor for ferric aerobactin [validated, MU
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-725/Product: cloacin receptor #status predicted <MAT>

Query Match 3.9%; Score 108; DB 2; Length 725;
Best Local Similarity 20.4%; Pred. No. 4;
Matches 79; Conservative 53; Mismatches 154; Indels 102; Gaps 14;

QY 80 LSPAYLRFGGTKDFLIFDPNKEPT--SEERSYQSDNNIDICGSERSVADVLKQM-- 135
D 196 LSVAYQKFGW-----FDNGDATLLDNTQTGLQYSDRLDIMGTGTLNIDESRQLQIT 249
QY 136 -----EWPFBELLLIREQYQ-----RE 152
D 250 QYYSQGGDDYGLNLGKFSAIRGTSTPFFVNSGNSDRIPGTERHLLISLQYSDFSFLGQE 309
QY 153 FKNSTYSRSVDMLYSFAKCSRDLIFGLNALLRTPDLRWNSNAQLLLNYCSKGNYS 212
D 310 LVGVVYRDESIRFYFPPTWNANKQVTAFFSSSQDQD---QYGMKLTLSKPMDWGQIT 365
QY 213 WEIGHNEPNSFWKAQISIDGLQGEDFVELHKLQKSAFONAKLYGPDICQPGKTVKLL 272
D 366 WGLDADHERFTS-----NOMFFDLAQSASGGLNKKIY--TTGRYPSYDITNL 412
QY 273 RFLKAGGEVIDSLTWHHYLNGRVATKEDFLSSDVLDTFLSVOKILKVTKEMTPGKV 332
D 413 AAFLOSQYDINILFT-----LNGGVRYQ---YTENKIDDFIGVAQO-----RQIGAK-- 457
QY 333 WLGETSSAYCGGAPLLSNTFAAGFM-----WLDKGLSLAQLGIEVVMRQVFFGAGN 383
D 458 --ATSADAFWLSRLRHLFLNAGLLMHIETPQAW-----LNFSSQVELPDPGKYGRGI 510
QY 384 YHLVDENFEPLDYLWLSLLFKKLVGPKV 411
D 511 YGAAVNGHLPLTKS--VNVSDSKLEGVKV 537
```

RESULT 8

```
A31842
endo-1,4-beta-xylanase (EC 3.2.1.8) Z precursor - Clostridium thermocellum
N:Alternate names: xylanase Z
C:Species: Clostridium thermocellum
C:Date: 31-Mar-1990 #sequence_revision 11-Apr-1997 #text_change 09-Jul-2004
C:Accession: A31842
J:Grepinet, O.; Chebrou, M.C.; Beguin, P.
R. Bacteriol. 170, 4582-4588, 1988
A:Title: Nucleotide sequence and deletion analysis of the xylanase gene (xynZ) of Clostri
A:Reference number: A31842; MUID:89008072; PMID:3139632
A:Accession: A31842
A:Molecule type: DNA
A:Residues: 1-837 <GRE>
A:Cross-references: UNIPROT:P10478; UNIPARC:UPI000013909C; GB:M22624; NID:gl44931; PIDN:f
C:Genetics:
A:Gene: xynZ
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A:Pathway: xylan degradation
C:Superfamily: Clostridium endo-1,4-beta-xylanase Z; Clostridium cellulase repeat homolog
C:Keywords: duplication; extracellular protein; glycosidase; heat-stable protein; hydroli
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-837/Product: endo-1,4-beta-xylanase #status predicted <MAT>
F:326-419/Domain: Clostridium xylanase A repeat homology <CXA>
F:430-453/Domain: Clostridium cellulase repeat homology <CCR1>
F:464-487/Domain: Clostridium cellulase repeat homology <CCR2>
F:548-834/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F:645,754/Active site: Glu #status predicted
```

Query Match 3.8%; Score 107; DB 1; Length 837;
Best Local Similarity 19.4%; Pred. No. 6;
Matches 86; Conservative 64; Mismatches 135; Indels 158; Gaps 23;

```
QY 14 LRALTQGTAGTAPTAKDVLDLEFYTKRLFQSVSFSLSI-TIDASLATDPRFLFGSPR 72
D 484 LRIITEPPGGDQVTN-----PSVTPTQTPIPTISGNALRD----- 520
QY 73 LRALAGLSPAYLRFGGTKDFLIFDPNKEPTSEERSYQSDNNIDICGSERSVADVLK 132
D 521 -YAEARGIKI-----GTCVNYFPYN-NSDPT--YNSILQREFSMVVCENE-MKFEDALQP 569
QY 133 LQMEWPFQELLRLREQYQREFKFNSTYSRSVDMLYSFAKSRDLIFGLNALLRTPDLRW 192
D 570 RQNVDF-----SKGDQLLAFERNMQ-----MRGHTLIW 600
QY 193 NSSNAQLLIN-----YCSSKGYNISWELGNE-----PNSFWK 224
D 601 HNQNPSWLITNGNWRNDSLLAVMKNHITVTMTHYKGIKIVWDVANECMDSDSGNGLRSSTWR 660
QY 225 KAQISIDGLQGEDFVELHKLQKSAFONAKLYG-----PDIGQPRGKTVKLLRSFLKAG 279
D 661 NV-----IGQYLDYAFRYAREADPDALLFYNDYNIEDLGPKNVAFNMVKS-MKER 711
QY 280 GEVIDSLTWHYILNGRVATKEDFLSSDVLDT-----FILSVQKI-LKVTKEMTPG 329
D 712 GVPIDGVGFQCHPING---MSPEYLAS--IDQNKYRAEIGVIVSFTIDIRIPOSENPA 766
QY 330 -----KKVWLGETSSAYCGGAPLLSNTFAAGFM--LDKUGLSAQLGIEVVM 374
D 767 TAFQVQANNYKELMKIKLANPN-----CNTFV---MWGPTDKY-----TWI 804
QY 375 RQVFFGAGNYHLVDENFEPLDPY 397
D 805 PGTFFPGYGNPLIYDSNYPKPAY 827
```

RESULT 9

```
T50113
N:Contains: 3-dehydroquininate dehydratase; 3-dehydroquininate synthase (EC 4.6.1.3)
C:Species: Schizosaccharomyces pombe
```

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: T50113
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
Submitted to the EMBL Data Library, February 2000
A;Reference number: Z25039
A;Accession: T50113
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1573 <SEE>
A;Cross-references: UNIPROT:Q9P7R0; UNIPARC:UPI0000125F22; EMBL:AL157734; PIDN:CAB75770.
A;Experimental source: strain 972h(-); cosmid c1834
C;Genetics:
A;Gene: SPDB:SPAC1834.02
A;Map position: 1
A;Superfamily: pentafunctional Arom protein; 3-dehydroquininate dehydratase homology; 3-dehydroquininate dehydratase homology
C;Keywords: carbon-oxygen lyase; phosphorus-oxygen lyase
F;407-835/Domain: 3-phosphohikimate 1-carboxyvinyltransferase homology <PSK>
F;1035-1279/Domain: 3-dehydroquininate dehydratase homology <DQD>

Query Match 3.8%; Score 106.5; DB 2; Length 1573;
Best Local Similarity 20.3%; Pred. No. 17;
Matches 115; Conservative 85; Mismatches 190; Indels 177; Gaps 30;

QY 19 QGTPAGTAPTKDVVDLEFYTKLQFQSPSPFLSITIDASLAT-----DP-RFLTFLGSPR 72
DB 699 QGPPKGLKPLESIDME-----TWTDAPLTASVVAACNVSEGDVPVTRITGIANQR 750

QY 73 LR-----ALARGSPAYLFGGKTDFLFDPN-KEPTSEERSYQSQDN----- 116
DB 751 VKECNRITAMVHELAKGCVTGELEDGIYIFGKYKELKPEEGIYYDDHRIAMSFSVL 810

QY 117 NDICGSEVSADVLRLKLOMFWPFOELLRLREYQREFKNSTYSSVDMLYSFAKCSRLD 176
DB 811 SLICPSRTLIID-KACVEKTPYWM-VDLHQSGFGVKLTGAT-SVASDPLKGSISKNSAII 867

QY 177 LIFGLNALLRTP-----DLRWNSSNAQLLLNCSSKGYNIWELGNENPSF 222
DB 868 LI-GMRGAGKTTICKITIAKQLNFKFLDL-----DELLEDYLEMPIAEVIFRMG----- 914

QY 223 WKKAQISIDGLQGEDFVELHKLQSAFQNAKLYGPDIGOPRGKTVK-----LLRSFL 276
DB 915 W-----DAFRLEE-----HKVLRFKTEHPEGY-----VAASGGVVIEMDESRLNLFV 959

QY 277 KAGGEVIDSLTHHYLYNGRVATKEDFLSSDVLDTFILSVQKILKVKEMTPGKKVWLGE 336
DB 960 KEGIVL-----HVHVN--LEHKVLSYSEDQTPYTKQDSIDDDYKR-----RHVVYRE 1007

QY 337 TSSAYGGGAPLLSNTFAAGFMWLDKGLSALQGLIEVVMRQVFFGAGNYHLVDENFEPLPD 396
DB 1008 CRSHY-FISPVLSN-----QVIDEKIQ----- 1028

QY 397 YWLSLFLPKLVGSKVLMRVKGPDRSKRLRVYLHCTNHYHPRYREGDLTYLVNLHNTVKH 456
DB 1029 YMSRFLDVTGSSQVLQKFKTKRSTF-----LTNYPRIEDALPTL-----RD 1073

QY 457 LKLPMPFSPRVDKYLLKPKFGSLKSVQLNQTLKMWDEQ-----TLPAL-TEK 507
DB 1074 VTGCDIAEVRVD-YLKDPKSSNGISS-----LDFVAEQISLLRCSCTTLPIITIR 1123

QY 508 PLPAG-----SSLSVPAPFSYG 523
DB 1124 TISQGGFLFPNDKEEAKELMSARYG 1150

RESULT 10
S75363
hypotheoretical protein s111913 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S75363
R;Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75363
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-599 <KAN>
A;Cross-references: UNIPROT:P73250; UNIPARC:UPI00000C0CB6; EMBL:D90904; GB:AB001339; NID: A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.8%; Score 106; DB 2; Length 599;
Best Local Similarity 19.6%; Pred. No. 4.3;
Matches 114; Conservative 79; Mismatches 220; Indels 168; Gaps 31;

QY 6 LLLW---LWGR-LRALTQGTAGTAPTKDVVDLEFYTKLQFQSPSPFLSITIDASLATD 61
DB 64 LAAMQTGLWGEFLRTLTVQPGG-----LDHWSVSLF-TFDPRLGSKIFADWAAE 112

QY 62 PRFLTFLGSP-RLRALARGLSPAYLRP-----GGTKT-DPLIFDPNKEPTSEER 108
DB 113 LPNLTWTIANQCPMEVVKNGDRLVGVRFPDYETIRARVILDTGTELGLDALLG-----DIGHR 167

QY 109 SYMQSQDNNDICGSEVSADVLRLKLOMFWPFOE---LILLREYQREFKNSTYSSVDM 165
DB 168 WGMQWQDKFD-----EPSCPIAPNEMTBEPYQSPPTWVFLLRKTTNNQ---TNIAEPTIDV 221

QY 166 LYSFAKCSRLDLIFGLNALLRTPDLRWNSSNAQLLLNCSSKG---YNISWELGNENPSFW 223
DB 222 AQDFTH-----TWQNYGEKDFLYGQLPGEHYMINWPIAG---NDYG 260

QY 224 KKAQISIDGLQGEDFVELHKLQSAFQNAKLYGPDIGOPRGKTVKRLSFLKAGGEVI 283
DB 261 KDLNRLLGGEKEKQTYL-----KEAQYSYAVAYLYQKHHSNLELATGIPPTGDIS 313

QY 284 DSLTWHHYLYNGR-----VATKEDFLSSDVLDTFILSVQKILKV-----T 323
DB 314 TAFALHPHYRESRLKGOQVITERDILPQGVASPLIYNQKVTSGVGNVANDHHYPGYE 373

QY 324 KEMTPGKKVWLGE-TSSAYGGGAPLLSNTFAAGFMWLDK----- 361
DB 374 PPLTPKSLIWGRWGTGPTTTPFPALLSTARGYLPCENKISVSHWANGSTRLOPLVMNT 433

QY 362 ---LGLSALQGLIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFLKLVGPKVLMRVKG 418
DB 434 QGAVGMIAALSVE-----KNCDP-QDLDVDVDQEALIG----- 465

QY 419 PDRSKRLRVYLHCTNHY--HPRYREGDLTYLVNLHNTVKHLKLPMPFSPRVDKYLLK-P 475
DB 466 -DRRAPAAVIFLNLVDPDHPERWQWQ--QYVLD--NPDQY----PPSGHCPVDVNLQALP 516

QY 476 FGSD-----GLLSKSVQLNGOTLKMWDEQT---LPALTEKP 508
DB 517 LSKSQSVYTGELQKS---EHQTYQLICQSQGKILKVITERP 554

RESULT 11
RNZ22
genome polyprotein - human respiratory syncytial virus (strain A2)
N;Alternate names: polymerase L protein
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: human respiratory syncytial virus
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004
C;Accession: A40317; A28319; F50048
R;Stec, D.S.; Hill III, M.G.; Collins, P.L.
Virolgy 183, 273-287, 1991
A;Title: Sequence analysis of the polymerase L gene of human respiratory syncytial virus
A;Reference number: A40317; MUID:9127486; PMID:2053282
A;Accession: A40317
A;Molecule type: mRNA
A;Residues: 1-2165 <STE>

A;Residues: 668-690, 'M', 692-753 <MAR>
A;Cross-references: UNIPARC:UPI0000176348; EMBL:X51541
R;Sotttrup-Jensen, L.; Sand, O.; Kristensen, L.; Fey, G.H.
J. Biol. Chem. 264, 15781-15789, 1989
A;Title: The alpha-macroglobulin bait region. Sequence diversity and localization of cleavage sites in the alpha-macroglobulin family.
A;Reference number: A34230; MUID:89380162; PMID:2476433
A;Accession: A34230
A;Molecule type: protein
A;Residues: 670-752, 'Q', 754-759 <SOT>
A;Cross-references: UNIPARC:UPI0000176349
C;Genetics:
A;Gene: GDB:PZP
A;Cross-references: GDB:120330; OMIM:176420
A;Map position: 12p13-12p12.2
A;Introns: 73/3; 116/2; 146/3; 169/3; 703/1; 753/2
A;Note: the list of introns may be incomplete
C;Superfamily: alpha-2-macroglobulin
F;685-735/Region: bait region

Query Match 3.7%; Score 104.5; DB 2; Length 1482;
Best Local Similarity 20.8%; Pred. No. 22;
Matches 113; Conservative 72; Mismatches 188; Indels 169; Gaps 29;

Qy 9 WLWGLRALTGQT-----PAGTAPTQDVVDLEFYTKRLFSQVSPSFLSIIT 54
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
745 WIWELVAVNSSGVAEVTVPDITETWKAGAFCLSEADAGLGISSTASURAFOPFFVELTM 804
Qy 55 DASLATDPRF-----LTFLG----SPRLRALARGLSPAYLRFGCTKTDFLIDPNKEP 103
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
805 PYSVIRGEVFTLKATVLNLYPKICRVSVQLKA-----SPAFLASQNTK----- 847
Qy 104 TSEERSYMQSDNDICGSERVSADVLRLQMFWFOELLLLREOYREFKNSTYSRSVV 163
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
848 --GEESYC-----ICGSRQT-----LSWTVT-----PKTLGNVNFS--VSA 880
Qy 164 DMLYFAKCSRLDLIFGLNALRTPLDRWNSSNAQLLL-----NYCSSKGYN 210
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
881 EAMQSLELCG-----NEWEVEPEIKRKDTVIKTLLEAVEAGEIEQKTFSSMTCASGAN 932
Qy 211 ISWELGNE--PNSFWK---KAQISIDGQLGEDFVELHKLQK---SAFONAKLYGPDIG- 262
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
933 VSEQSLKLPSNVVKSESARASFVGILDILSAMQNIQNLLQMPYCGGQQNWLPAPNIYV 992
Qy 263 -QPRGKTVKLLRSF-LKAGEVID----SLTWHHYYLNGRVAT-----KEDFLSSD 307
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
993 LNYLNETOQLTQBIKAKAVGYLITGYORQLNYKHQ--DGSYSTFGERYGRNQGTWLTA 1050
Qy 308 VLDTFILSVOKIL-----KVTKEMTPGKKWLGE---TSSAYGGGAPILLSNTFAAGFMWLD 360
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
1051 VLKTEFAQARSYFIIDEAHITQSILT-----WLSOMQKONGCFRSSGSLNNAIKGGVE--D 1103
Qy 361 KLGLSNQLGIEV-----VMRQVVF-----GAGNYHLVDENPERLPDYWL 399
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
1104 EATLSAYVTIALLBEIPLVTNPVRNALFCLESAMNVAKEGTHGSHVYT---KALLAYAF 1160
Qy 400 SLLFKKLGVKPKVLMRSVK-----GPDRSKLRV-YLHCTTNVYHPRYREGDLTIY 446
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
1161 SILGKQNRNEIILNSLDKEAVKEDNLVHWERPQRKPAPVGHLYQTQA---PSAEVENTSY 1217
Qy 447 VL 448
Db ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
1218 VL 1219

RESULT 14
S61166
N;Alternate names: YDR371w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: S61166
Submitted to the EMBL Data Library, June 1995
R;Ding, H.

A;Description: The sequence of *S. cerevisiae* coamid 9481.
A;Reference number: S61159
A;Accession: S61166
A;Molecule type: DNA
A;Residues: 1-511 <DIN>
A;Cross-references: UNIPROT:Q06350; UNIPARC:UPI000006A221; EMBL:U28373; NID:g849184; PID:
A;Experimental source: strain S288C (AB972)
C;Genetics:
A;Gene: MIPS:YDR371w
A;Cross-references: SGD:S0002779
A;Map position: 4R
C;Superfamily: Serriatia marcescens chitinase
C;Keywords: transmembrane protein
F;18-34/Domain: transmembrane #status predicted <TMM>

Query Match 3.7%; Score 104; DB 2; Length 511;
Best Local Similarity 19.8%; Pred. No. 4.9;
Matches 72; Conservative 52; Mismatches 132; Indels 108; Gaps 15;

Qy 200 LLYCCSKGYNISWELNPNPFWKQAQISIDGL-----QLGEDFVELHKL 246
Db 165 LKNTCSDKPKVIMSIGWSDSNFKIIIKODKLLQNFDSSVETMFLGPDIDL---- 220
Qy 247 QKSAFQNAKLYGPDIGOPRG--KTVKLLRSFLKA-GGEVIDSLTWHHYYLNGRVATKEDF 303
Db 221 -----DWEPFGNNESEPRGVYLKVRMLRLKLNLSLEQIFGKRTEHDHFOLSTAAPAFKOK 274
Qy 304 LSSDLDTFILSVOKILVKTKEMTPGKKVWLGETSSAYGGGAPLLSNT-----FAAGPWM 358
Db 275 L-----FVLPIEIQDYVDYNNMTYDYGSMSETTYGHSNLFSDELNGNFAMHYW- 326
Qy 359 LDKLGLSAQIGIEVVMRQVPFGAGNYHLVDENFEPLPD--YWLSLLFKLVCPKVLMRSV 416
Db 327 IDRFVNSR---KLVLGNMAAYGR-SPHIKDNKFEPNQNTVLINKIFKGVGKPTKEIDKA 382
Qy 417 KGPD-----RSLRVYLH---- 429
Db 383 DGKEGIWPYNLPKIGTIEQDPKYVSAYCFDEKNISIFISYDNTKSVTKAEYVTHNLG 442
Qy 430 -----CTNVYHPRYREGDLTLVNLHNVTQHKL--PPPMFSR-PVDKYLKLPFGSD 479
Db 443 GGFWESCGEAYANESRS-----LINAFNEGHLFNVSCKPSIFQDVVRVKYYLKNKYGDG 496
Qy 480 GLLS 483
Db 497 GFLS 500

RESULT 15
T44483
receptor-like protein iuta [imported] - Shigella flexneri
C;Species: Shigella flexneri
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44483
R;Moss, J.B.; Cardozo, T.J.; Zychlinsky, A.; Groisman, B.A.
Mol. Microbiol. 33, 74-83, 1999
A;Title: The selc-associated SHI-2 pathogenicity island of Shigella flexneri.
A;Reference number: 222779; PMID:99340540; PMID:10411725
A;Accession: T44483
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-732 <MOS>
A;Cross-references: UNIPROT:Q9XCH0; UNIPARC:UPI00000BE41C; EMBL:AF141323; NID:g5532445; I:
A;Experimental source: strain M90T; serotype 5a
C;Genetics:
A;Gene: iuta

Query Match 3.7%; Score 104; DB 2; Length 732;
Best Local Similarity 19.2%; Pred. No. 8.4;
Matches 83; Conservative 54; Mismatches 168; Indels 128; Gaps 15;

Qy 80 LSPAYLRFGGTKTDFLIPDPNKPTSERSTWQSQDNH--DICGSERVADVLKLOM-- 135

```
Db 195 LSVAYOKFGGW-----FDGNGDATLLDNTQTGLQHSNRLDIMGTGTLNIDESRQLQIIT 248
QY 136 -----EWPFOELLILREQYQ-----RE 152
Db 249 QYYKSGDDNYGLNLGKGFSAISGSSTPYVSKGLNSDRIPTERHILSLQYSDSDFLGQE 308
QY 153 FKNSTYSRSSVDMLYSPFAKCSRDLDFGLNALLRTPDLRWNSSNAQALLNYCCKGYNIS 212
Db 309 LVGQVYVRDESLRYPPFTVNANKQATAFSSSQDQTD---QYGMKLTLSQLMDGWQIIT 364
QY 213 WELGNEPNSFWKKAQISIDQLGDEDFVELHKLLOKSAFONAKLYGPDIGOPRGKTVKLL 272
Db 365 WGLDAHERFTS-----NQMFDLAQASAGGLNNHKIY--TTGRYPSYDITNL 411
QY 273 RSFLKAGGEVIDSLT---WHHYLNGRVATKEDF-----LSSDVL----- 309
Db 412 AAFLOSSYDINDIFTVSGGVRYQYTENRV---DDFDITYTQQOKIAAGKAISADAIPGGSV 468
QY 310 --DTFILSVOKILKVTKEM-----TFGKKVWIGETSSAYGGGAPLLSNTFAA 354
Db 469 DYDNFLFNAGLLMHITERQQALENFNSQGVALPDPGKYGRGIYCAAVNGHPLTKSV--- 525
QY 355 GFMWLDKGLLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSLLFYKLVGPVULMS 414
Db 526 -----NVSDSKLEGVKVDSYEL-----GWRFIGDNLRTQIAAYYLSLNKSVERNKDLTI 574
QY 415 RVKGPDRSKLRVY 427
Db 575 SVKDDRR---RIY 584
```

Search completed: June 5, 2006, 12:21:46
Job time : 20.8454 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 132.991 Seconds
(without alignments)
3728.138 Million cell updates/sec

Title: US-10-645-659A-3
Perfect score: 2800
Sequence: 1 MLAPLLLLLWGLRALTGQ.....VPAFSYGFVIRNAKIAICI 536

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2783	99.4	536	1	HPSE_RAT
2	2576.5	92.0	535	1	HPSE_MOUSE
3	2336	83.4	574	2	Q333X8_SPALAX_GOLA
4	2336	83.4	574	2	Q333X9_SPALAX_GALI
5	2331	83.2	574	2	Q333X7_SPALAX_CARM
6	2323	83.0	574	2	Q333X6_SPALAX_JUDA
7	2218	79.2	558	1	Q333X5_SPAJD
8	2151	76.8	545	1	HPSE_BOVIN
9	2126	75.9	543	1	HPSE_HUMAN
10	1577	56.3	523	1	HPSE_CHICK
11	1290	46.1	533	2	Q4SYF6_TETNG
12	1171	41.8	592	1	HPSE2_HUMAN
13	1171	41.8	592	2	Q2MIH9_HUMAN
14	1048	37.4	597	2	Q4TB80_TETNG
15	725.5	25.9	255	2	Q4TG88_TETNG
16	669	23.9	515	2	Q8T108_BOMMO
17	410	14.6	543	1	HPSE3_ARATH
18	385	13.8	559	2	Q89F99_BRAJA
19	374	13.4	527	2	Q9LR08_SCUBA
20	368.5	13.2	526	2	Q5SNA6_ORYSA
21	368.5	13.2	541	2	Q69116_ORYSA
22	357.5	12.8	539	2	Q2QN56_ORYSA
23	353	12.6	536	1	HPSE3_ARATH
24	352.5	12.6	537	2	Q70YJ3_HORVU
25	350	12.5	529	2	Q6ZJE2_ORYSA
26	336	12.0	401	2	Q30324_ARATH
27	330.5	11.8	539	1	HPSE2_ARATH
28	330	11.8	516	2	Q447R5_SOLUS
29	299	10.7	537	2	Q43S03_SOLUS
30	272	9.7	506	2	Q37Q70_SPHAR
31	239.5	8.6	382	2	Q3E8P7_ARABIDOPSIS

32	170	6.1	1128	2	Q5TT65_ANOGA
33	154	5.5	510	2	Q2U0T3_ASPOR
34	139.5	5.0	935	2	Q9VE79_DROME
35	137.5	4.9	1053	2	P71329_FIBSU
36	137.5	4.9	1053	2	P77865_FIBSU
37	132.5	4.7	670	2	Q3JTG0_BURP1
38	131.5	4.7	463	2	Q6J797_BURPS
39	128	4.6	559	2	Q7SFB0_NEUCR
40	125.5	4.5	536	2	Q2UDS9_ASPOR
41	125	4.5	1175	2	Q5KTI5_9ALTE
42	124	4.4	739	2	Q8EWI3_MYCPE
43	123	4.4	689	2	Q5VNX0_NOCFA
44	122.5	4.4	628	2	Q2ZEV8_CALSA
45	120	4.3	721	2	Q3XV07_9PROT

ALIGNMENTS

RESULT 1

HPSE_RAT	ID	HPSE_RAT	STANDARD;	PRT;	536 AA.
AC	Q71RP1; Q9QZF8;				
DT	11-OCT-2005, integrated into UniProtKB/Swiss-Prot.				
DT	05-JUL-2004, sequence version 1.				
DT	07-MAR-2006, entry version 11.				
DE	Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:				
DE	Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].				
GN	Name=Hpse; Synonyms=Hep;				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;				
OC	Muridea; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE [MRNA].				
RC	Tissue=Placenta;				
RX	MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;				
RA	Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,				
RA	Parish C.R.;				
RT	"Cloning of mammalian heparanase, an important enzyme in tumor				
RT	invasion and metastasis.";				
RL	Nat. Med. 5:803-809(1999).				
RL	[2]				
RP	NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.				
RX	MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;				
RA	Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;				
RT	"Characterization of heparanase from a rat parathyroid cell line.";				
RL	J. Biol. Chem. 277:32459-32465(2002).				
CC	-!- FUNCTION: Endoglycosidase which is a cell surface and				
CC	extracellular matrix-degrading enzyme. Cleaves heparan sulfate				
CC	proteoglycans (HSPGs) into heparan sulfate side chains and core				
CC	proteoglycans. Also implicated in the extravasation of leukocytes				
CC	and tumor cell lines. Contributes to metastasis and angiogenesis				
CC	(By similarity).				
CC	-!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower				
CC	extent, by heparin and sulfamin (By similarity). Activated by				
CC	calcium and magnesium. Inhibited by EDTA.				
CC	-!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa				
CC	subunits, the proteolytic products (By similarity).				
CC	-!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.				
CC	Secreted, internalised and transferred to late endosomes/lysosomes				
CC	as a proheparanase. In lysosomes, it is processed into the active				
CC	form, the heparanase. The uptake or internalisation of				
CC	proheparanase is mediated by HSPGs. Heparin appears to be a				
CC	competitor and retain proheparanase in the extracellular medium				
CC	(By similarity).				
CC	-!- PTM: Proteolytically processed. The cleavage of the 65 kDa form				
CC	leads to the generation of a linker peptide, 8 kDa and 50 kDa				
CC	product. The active form, the 8/50 kDa heterodimer, is resistant				
CC	to degradation. Complete removal of the linker peptide appears to				
CC	be a prerequisite to the complete activation of the enzyme (By				
CC	similarity).				

Q5TT65	anopheles g
Q2U0T3	aspergillus
Q9VE79	drosophila
P71329	fibrobacter
P77865	fibrobacter
Q3JTG0	burkholderi
Q6J797	burkholderi
Q7SFB0	neurospora
Q2UDS9	aspergillus
Q5KTI5	microbulbif
Q8EWI3	mycoplasma
Q5VNX0	nocardia fa
Q2ZEV8	caldicellul
Q3XV07	magnetococ

CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF359508; AAQ15189.1; -; mRNA.
DR EMBL; AF184967; AAF04563.1; -; mRNA.
DR RGD; 61969; Hpsa.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; I.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
FT SIGNAL.
FT CHAIN 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 Linker peptide (By similarity).
FT CHAIN 151 536 /FTid=PRO.0000042267.
FT REGION 151 155 Heparanase 50 kDa subunit.
FT REGION 263 273 /FTid=PRO.0000042268.
FT ACT_SITE 218 218 Heparin/HS-binding (By similarity).
FT ACT_SITE 336 336 Heparin/HS-binding (By similarity).
FT CARBOHYD 155 155 Nucleophile (Potential).
FT CARBOHYD 193 193 N-linked (GLNAC. .) (By similarity).
FT CARBOHYD 210 210 N-linked (GLNAC. .) (By similarity).
FT CARBOHYD 452 452 N-linked (GLNAC. .) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 99.4%; Score 2783; DB 1; Length 536;
Best Local Similarity 99.4%; Pred. No. 9.8e-200;
Matches 533; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLRPLLLLMWGLRALTQCTPGTAGTPTKDVVLEFFYTKRFLFQSVSPFLSITIDASLAT 60
DB 1 MLRPLLLLMWGLRGALTQCTPGTAGTPTKDVVLEFFYTKRFLFQSVSPFLSITIDASLAT 60

QY 61 DRPFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLFPDNPKEPTSEERSYQSDNNNDIC 120
DB 61 DRPFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLFPDNPKEPTSEERSYQSDNNNDIC 120

QY 121 GSERVSADVLRLKQMEWPFQELLRLREYQREPKNSTYRSRSDMLYSPFAKCSRLDLIFG 180
DB 121 GSERVSADVLRLKQMEWPFQELLRLREYQREPKNSTYRSRSDMLYSPFAKCSRLDLIFG 180

QY 181 LNALRLTPDLRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
DB 181 LNALRLTPDLRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDFV 240

QY 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTHWHYLYNGRVATK 300
DB 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTHWHYLYNGRVATK 300

QY 301 EDPLSSDVLDTFLSVQKILKVTKEPMPGKVMWLGSETSSAYGGAPLLSNTFAAGFMWLD 360
DB 301 EDPLSSDVLDTFLSVQKILKVTKEPMPGKVMWLGSETSSAYGGAPLLSNTFAAGFMWLD 360

QY 361 KLGLSQAQLGIEVVMRQVFFGAGNYHLVDENFELPDYWLSSLKPLGPKVLSRVKGP 420
DB 361 KLGLSQAQLGIEVVMRQVFFGAGNYHLVDENFELPDYWLSSLKPLGPKVLSRVKGP 420

QY 421 RSKLRVYLCTNYHPRYREGDITLVYLNHNTYTKHLKLPMPFSRPVDKYLKPFSGSDG 480
DB 421 RSKLRVYLCTNYHPRYREGDITLVYLNHNTYTKHLKLPMPFSRPVDKYLKPFSGSDG 480

QY 481 LLKSKVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASGYGFFVIRNAKIAACI 536
DB 481 LLKSKVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASGYGFFVIRNAKIAACI 536

RESULT 2
HPSE_MOUSE STANDARD; PRT; 535 AA.
AC Q6YGI; O8K3K3;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (BC 3.2.-.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=Hpsa; Synonyms=Hpa;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=SJL/J; TISSUE=Spleen;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RT Nat. Med. 5:803-809(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND
RP SUBUNITS.
RC STRAIN=FVB; TISSUE=Embryo;
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
RT "Cloning, expression, and purification of mouse heparanase";
RL Protein Expr. Purif. 26:425-431(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
RA Gong F., Jemth P., Galvis M.L.E., Vlodevsky I., Horner A., Lindahl U.,
RA Li J.-P.;
RT "Processing of macromolecular heparin by heparanase";
RL J. Biol. Chem. 278:35152-35158(2003).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hilt D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jaki T., Kanapin E., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lazarev L.F., Lazarevic D., Lipovich L., Liu J.,
RA Luoni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Motaghi-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K.,

RA	NUCLEOTIDE SEQUENCE.
RP	TISSUE=Kidney;
RC	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviwi A.;
RA	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RT	Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RL	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL; AM085491; CAJ30018.1; --; mRNA
SQ	SEQUENCE 574 AA; 64555 MW; 48EBEFEC7D0BCB34 CRC64;
	Query Match 83.4%; Score 2336; DB 2; Length 574;
	Best Local Similarity 83.0%; Pred. No. 3.5e-166;
	Matches 445; Conservative 40; Mismatches 49; Indels 2; Gaps 1;
Qy	1 MLRPLLLWGRURALVTQGTPTAGTAPTKOVVDLEFFYTKRLFQSVSPFLSITIDASLAT 60
Db	41 MLRLSLLLWGPLSPVQCILA--QAEDVVELEFSTQRPDLHLVSPFLSITIDANLAT 98
Qy	61 DPRELTGLSPRLALARGLSPAYLRGGTKTDFLI FDPNKEPTSEERSYQSQNDNDIC 120
Db	99 DPRELTGLSPKRLALARGLSPAYLRGGTKTDFLI FDPKKEPSHERSYKQSVNHIDIC 158
Qy	121 GSERVSADVLRLKQWEPFPQELLRLREQYQREFKNSTYSRSSVDMLYFAKCSRLDLIFG 180
Db	159 RSGAIPAVVVRRLQVEWFPQELLRLREQYQKEFKNSTYSRSSVDMLYTFARCSGLDLIFG 218
Qy	181 LNALLRTPDLRWNSNAQLLNNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
Db	219 LNALLRTADFRWNSNAQLLNNYCSSKKNYDISWELGNEPNSFWKKAHTSIDGLQGEDYI 278
Qy	241 ELHKLLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRVATK 300
Db	279 ELHKLLRKSTLKNVLYGPDVGQPRGKTVKLLRSFLKAGGEVIDSVTWHHYLLNGRIATK 338
Qy	301 EDFLSSDVLDTFILSVOKILAVTKEMTPGKKVWLGETSSAYGGCAPLLSNTFAAGFWMLD 360
Db	339 EDFLSPDVLDTFILSVOKILQVSETRPGKKVWLGETSSAYGGCAPLLSNTFAAGFWMLD 398
Qy	361 KLGLSAQLGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFLFKLYGPKVIMSRVKGPD 420
Db	399 KLGLSSQMGIEVVMRQVFFGAGNTHLVNDKNEPELPDYWLSSLFLFKLYGSKVLMARVKGPD 458
Qy	421 RSKURVYLHCTNVTHPRYREGDLTYLVNLNHTVTKHLKLPPEMFSRPVDKYLKLPFGSDG 480
Db	459 RSKLRVYLHCTNINHPRYQEGDLTYALNLYNVTKHLKLPQLFENKPDYKYLKLPFGPGG 518
Qy	481 LLSKSVOLNGTTLQWVDEOTLPALTEKPLAGSSISVPAPSYGFFVIRNAKIAACI 536
Db	519 LLSKSVQLNGQALKWVDQDTLPALTEKPLRPGSSGLGPAFSYGFFVIRNAKAAACL 574

RESULT 4	
Q333X9	PRODE PRELIMINARY; PRT; 574 AA.
ID	Q333X9_9RODE
AC	Q333X9_9
DT	06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT	06-DEC-2005, sequence version 1.
DT	07-FEB-2006, entry version 3.
DE	Heparanase.
GN	Name=hpa;
OS	Spalax galili.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Spalacidae; Spalacinae; Spalax.
OX	NCBI_TaxID=164323;
FP	(1)
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviwi A.;

"adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene cloning and identification of a novel splice variant.",
Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
(2)
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Kidney;
RC Naesser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RA "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
DD
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match 83.4%; Score 2336; DB 2; Length 574;
Best Local Similarity 83.0%; Pred. No. 3.5e-166;
Matches 445; Conservative 40; Mismatches 49; Indels 2; Gaps 1

Qy 1 MLRLPLLWLRGRALRTQGT*PAGTAPT*QDVVDLEFYTKRLFQSPSPFLSITIDASLAT 60
Db 41 MLRLSLLWLRGPLSPVQCTLA--AQAEVVLEFFSTQRP*HLVSPFLSITIDANLAT 98
Qy 61 DPRFLT*LGSPRLRALAGLSPAYLRFGGKT*DFLIFDPNKEPTSEERSYMQSDNNDC 120
Db 99 DPRFLT*LGSPKRLRALAGLSPAYLRFGGKT*DFLIFDPKKEPSHEERSYMQSVNHDIC 158
Qy 121 GSERVSADVLRLKQWEPFQELLRLRLROYQREFKNSTYRSRVDMLYSFAKCSRLDLIFG 180
Db 159 RSGAIPAVVVRRLQVWEPFQELLRLREQYQDFKNSTYRSRVDMLYTFARCSGLDLIFG 218
Qy 181 LNALRLTPDLRWNSNAQLLNNYCSSKGYNTISWELGNPNPSFWKKAQISIDGLQLGEDPV 240
Db 219 LNALRLTADPRWNSNAQLLNNYCSSKNYDISWELGNPNPSFWKKAHISIDGLQGEDYI 278
Qy 241 ELHKLLKSAFONAKLYGPDIGQPRGKT*VKLLRSFLKAGGEVIDSLTWHYYLYNGRVATK 300
Db 279 ELHKLLRKSTLKNVKLYGPDVGQPRGKT*VKLLRSFLKAGGEVIDSVTWHYYLYNGRIATK 338
Qy 301 EDFLSSVDLTFILSVQKILKVT*EMTPGKKWIGETSSAYGGGAPLISNTFAAGFMWLD 360
Db 339 EDFLSPVLDLTFILSVQKILQV*EETPRGKKWIGETSSAYGGGAPLISNTFAAGFMWLD 398
Qy 361 KIGLSAQGLGVVWVRQVFFGAGN*HLVDENPELPDYWLSLLFKLVGP*KVLMRSVKGPD 420
Db 399 KIGLSAQMGLEVVWVRQVFFGAGN*HLVDKNPELPDYWLSLLFKLVGSKVLMARVKGP 458
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTLYVLNLHN*TKHLKLP*PPMFSPRVDKYLKLPFGSDG 480
Db 459 RSKLRVYLHCTNINHPRYQEGDLTLYALNLYN*TKHLKLPQLFKNKPDKYLKLPGLPGG 518
Qy 481 LLSKSVQLNGQTLK*WDBQ*TL*PALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536
Db 519 LLSKSVQLNGQAL*WVDD*QTL*PALTEKPLRPGSSILGLPAFSYGFVIRNAKVAACL 574

RESULT 5
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
ID AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Murioidea; Spalacidae; Spalacinae; Spalax.
NCBI_TaxID=164324;
RN [1]

RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avioli A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL: AM085492; CAJ30019.1; -, mRNA.
SQ	SEQUENCE 574 AA; 64459 MW; 9F1D19DCBADD99DE CRC64;
	Query Match 83.2%; Score 2331; DB 2; Length 574;
	Best Local Similarity 83.0%; Pred. No. 8.4e-166;
	Matches 445; Conservative 39; Mismatches 50; Indels 2; Gaps 1;
QY	1 MLRPLLLLLWGLRALTLTGTPAGTAPTKOVVDLEFVTKRLFQSVSPSFLSITIDASL 60
DB	41 MLRLSLLLLWGLPSLPLVQCILA--AAQEDVVELEFSTQRPDLHLVSPFLSITIDANL 98
QY	61 DPRFLTPLGSPRLRALAGLSPAVLRFGGKTDFLIIDPNKEPTSEERSYQSQDNNDIC 120
DB	99 DPRFLTPLGSPKRLALAGLSPAVLRFGGKTDFLIIDPKPEPSHEERSYWKSVQNHDI 158
QY	121 GSERVSADVLRLKQMEWPPQELLRLRLRQYQREFKNSTYSRSSVDMLYSFAKCSRLDLIF 180
DB	159 RSGAIPAVVVRRLQVWPFOQLLLRQYQKEFKNSTYSRSSVDMLYTFARCSGLDLIF 218
QY	181 LNALRRTPDLRWSSNAQLLNNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
DB	219 LNALRRTPDLRWSSNAQLLNNYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDYI 278
QY	241 ELAKLLQKSAFONAKLYGPDIGPQGRKTVKLLRSFLKAGGVIDSLTWHYYLNGRVATK 300
DB	279 ELRLKRLKSTLKNVKLYGPDVGQPRGKTVKLLRSFLKAGGVIDSVTWHYYLNGRIATK 338
QY	301 EDFLSPLVDLTFILSVQKILKVTXKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFWMLD 360
DB	339 EDFLSPLVDLTFILSVQKILQVVEETPRGKKVWLGETSSAYGGAPLLSNTFAAGFWMLD 398
QY	361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLFKLYGPKVLMRSRVKGP 420
DB	399 KLGLSAQMGIEVVMRQVFFGAGNYHLVDKNPEPLPDYWLSSLFKLYGSKVLMARVKGPD 458
QY	421 RSKLRVYLHCTNVYHPRVREGDLTLVYLNLNHNVTKHLKLPPEMSPRPVDKYLKLPFGSDG 480
DB	459 RSKLRVYLHCTNVHNPYQEGDLTLVYALNLYNVTKHLKLPQLFNKVPDKYLKLPFGPG 518
QY	481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFSGYFFVIRNAKIAACI 536
DB	519 LLSKSVQLNGQALKMWDDQTLPALTEKPLPGSSLSLPAFSGYFFVIRNAKVAACL 574
RESULT 6	
Q333X6	SPAJD
ID	Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC	Q333X6
CC	Q333X6
DT	06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT	06-DEC-2005, sequence version 1.
DT	07-FEB-2006, entry version 3.
DE	Heparanase.
GN	Name:hpa.
OS	Spalax judaei (Blind subterranean mole rat).
OC	Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Spalacidae; Spalacinae; Spalax.
NCBI	TaxID=134510;
RN	[1]
RN	NCBIOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avioli A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene

RT	Cloning and identification of a novel splice variant.";
RL	[Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005)].
RN	[2]
RP	NUCLEOTIDE SEQUENCE.
RC	TISSUE=Kidney;
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT	cloning and identification of a novel splice variant.";
RL	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
SQ	EMBL; AM085493; CAJ30020.1; -; mRNA.
DR	SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;
Query Match	83.0%; Score 2323; DB 2; Length 574;
Best Local Similarity	82.8%; Pred. No. 3.3e-165;
Matches 444,	Conservative 39; Mismatches 51; Indels 2; Gaps 1
Qy	1 MLRPLLALLMWGRURALTGQTPTAGTAPTKVDVLDLFEYTKRLFSQSPSLSTIDASLAT 60
Db	41 MLRLSLLLMWGLPSLVQCILA--AQAEDEVVEFEFTQRPLHLVSPSPSLSTIDANLAT 98
Qy	61 DPRELTLFGSPRLRALARGLSPAYLRFGGTTKDFLI FDPNKPEPTSEERSYWKOSQDNNDIC 120
Db	99 DPRELTFLGSPKRLALARGLSPAYLRFGGTTKDFLI FDPKKEPSEHEERSYWKSNVHDCI 158
Qy	121 GSERVSADVLRLKLQEWEPFFQELLRLLEQYQREFKNSTYSRSSVDMLYSPFAKCSRLDLIFG 180
Db	159 RSGNAIPAVVRRLQEVFPFQELRLLEQYQEFKNSTYSRSSVDMLYTFARCSGLDLIFG 218
Qy	181 LNALLRTPDRLWNSSNAQLLNYYCSSKGYNITSWELGNENPFNSFWKKAQISIDGLQGEDVF 240
Db	219 LNALLRADFRWNSSNAQLLNYYCSSKNYDISWELGNENPFNSFWKKAHISIDGLQGEDVI 278
Qy	241 ELHKLLOKSAFONAKLIGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYYLNGRVATK 300
Db	279 ELRKLLKKSTLKNVKYLGPDVGPGRGKTVKLLRSFLKAGGEVIDSVTHHHYYLNGRIATK 338
Qy	301 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVMILGETSSAYGGGAPLLSNTFAAGFMWLD 360
Db	339 EDFLSPDVLDTFILSVQKILQVZETRGKKVMILGETSSAYGGGAPLLSNTFAAGFMWLD 398
Qy	361 KLGLSAQLGIEGVNMRQVFFGAGNYHLVDENPELPDYWLSSLFKKLVGPKVMSRVKGGPD 420
Db	399 KLGLSAQMGIEGVNMRQVFFGAGNYHLVDKNFEPLPDYWLSSLFKKLVGSKVLMARVKGGPD 458
Qy	421 RSKURVYLHCTNVTHPYREGDGLTYLVNLHNVTKHKLPPPMFSRPVDKYLLKPGGSDG 480
Db	459 RSKURVYLHCTNIHPHYREQEDGLTYALNLVNVTKHKLPLYQLFNKPVDKYLLVIPLPGGG 518
Qy	481 LLKSXVOLNGGTQKWVDQETLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536
Db	519 LLKSXVOLNQALKXWDQTLPALTEKPLRPGSSILGLPAFSYGFFVIRNAKVAACL 574
RESULT 7	
Q333XS SPAJD	PRELIMINARY; PRT; 558 AA.
ID Q333XS SPAJD	
AC Q333XS;	
DT 06-DEC-2005,	integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005,	sequence version 1.
DT 07-FEB-2006,	entry version 3.
DE Heparanase.	
GN Name=hpa;	
OS Spalax judaei (Blind subterranean mole rat).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	
OC Muridea; Spalacidae; Spalacinae; Spalax.	
OX NCBI_TaxID=134510;	
RN [1]	
RP NUCLEOTIDE SEQUENCE	

Db 69 TDPFRFTFLGSSKLTTLARGLAPAYLRFNGKNGDFLIFDPKKEPAFEERSYWLQSNDI 128
Qy 120 CGSERVSADVLRLKQWEPQOELLRLREOYQREBNSTYSRSSYDMLYSPAKCSRDLIF 179
Db 129 CKSGSPSDVEEKLRLWEPQOELLRLREOYQREBNSTYSRSSYDMLYTFASCGLNLF 188
Qy 180 GLNALLRTDPLRNSSNAQLLNLCSSGKYNISWELGNEPNSFWKKAQISIDGLQGLGDF 239
Db 189 GVNALLRTDTHWDSSNAQLLNLCSSKNTNISWELGNEPNSFORAKGIFINGQLGDF 248
Qy 240 VELHLKLQKSAFQNAKLIGPDIGOPRGKTIVKLRSFLKAGGEVIDSLTWHYYLNGRVAT 299
Db 249 IEFKLLGSAFNAKLIGPDIGOPRENTVKMLKSLFKAGGEVIDSVTWHYYNGRIAT 308
Qy 300 KEDFLSDVLDITFLSVQKILKVTYKWTMPGKVKWLGTSSTAYGAGBAPLLSNTFAAGFWL 359
Db 309 KEDFLNPDIIDTISVQKTLRIVEKIRPLKVKWLGTSSTAYGAGBAPLLSNTFAAGFWL 368
Qy 360 DKLGSLAQGLIEVVMROVFFGAGNYHLVDENFELPDYWLSSLFELKLVGPKVMSRVKGP 419
Db 369 DKLGSLARMGIEVVMROVFFGAGNYHLVDGNFELPDYWLSSLFELKLVGPKVMSRVKGP 428
Qy 420 DRSKRLVYLCTNVYHPRYREGDITLVNLHNVTKHLKLPMPFSPVDKYLKLPFGSD 479
Db 429 DRSKFRVYLCTNWKHPRYREGDITLVNLHNVTKHLKLPMPFSPVDKYLKLPFGSD 488
Qy 480 GLLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLVAPAFSGFFVIRNAKIAACI 536
Db 489 GLLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLVAPAFSGFFVIRNAKIAACI 545

RESULT 9

HPSE HUMAN
ID HPSE HUMAN STANDARD; PRT; 543 AA.
AC Q9Y2E1; Q53GE5; Q9UL39;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1999, sequence version 1.
DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).
GN Names=HPSE; Synonyms=HBP, HPA, HPA1, HPR1, HPSE1, HSE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=99353379; PubMed=10405343; DOI=10.1006/bbr.1999.0962;
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and functional expression of a human heparanase gene.";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.
RC TISSUE=Embryonic fibroblast;
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and expression.";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
RX PubMed=10395325; DOI=10.1038/10518;
RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Azmon R., Ishai-Michaeli R., Bican M., Pappo O., Peretz T., Michal I., Spector L., Pecker I.;
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";

Nat. Med. 5:793-802(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 466-468 AND 478-483.
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";
RN Nat. Med. 5:803-809(1999).
RN [5]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular damage.";
RL Glycobiology 10:467-475(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
RX PubMed=11547900; DOI=10.1023/A:1011375624902;
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O., Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
RT "Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis.";
RN J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
RN [7]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163, SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
RC TISSUE=Placenta;
RX PubMed=12713442; DOI=10.1042/BJ20000318;
RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R., Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S., Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M., Parekh R., Stubberfield C.;
RT "Biochemical characterization of the active heterodimer form of human heparanase (Hpal) protein expressed in insect cells.";
RL Biochem. J. 373:423-435(2003).
RN [8]
RP NUCLEOTIDE SEQUENCE [MRNA].
RA Pinhal M.A., Semedo P.;
RT "Cloned heparanase from MCF-7 cells.";
RN Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
RN [9]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Small intestine;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.;
RN Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
RN [10]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [11]
RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
RX PubMed=11123890; DOI=10.1021/bi002080p;
RA Hulet M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
RA Gready J.E., Parish C.R.;
RT "Identification of active-site residues of the pro-metastatic
RT endoglycosidase heparanase.";
RL Biochemistry 39:15659-15667(2000).
RN [12]
RN N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
RN ASN-217; ASN-238 AND ASN-459.
RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
RA Simizu S., Ishida K., Wierzbza M.K., Osada H.;
RT "Secretion of heparanase protein is regulated by glycosylation in
RT human tumor cell lines.";
RL J. Biol. Chem. 279:2697-2703(2004).
RN [13]
RN SUBCELLULAR LOCATION.
RX PubMed=15292202; DOI=10.1074/jbc.M402131200;
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodaysky I., Ilan N.;
RT "Heparanase uptake is mediated by cell membrane heparan sulfate
RT proteoglycans.";
RL J. Biol. Chem. 279:44084-44092(2004).
RN [14]
RN BIOPHYSICO-CHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
RA Cohen E., Atzmon R., Vlodaysky I., Ilan N.;
RT "Heparanase processing by lysosomal/endosomal protein preparation.";
RL FEBS Lett. 579:2334-2338(2005).
RN [15]
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
RA Abboud-Jarroos G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
RA Elgavish S., Peretz T., Vlodaysky I.;
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
RT human proheparanase.";
RL J. Biol. Chem. 280:13568-13575(2005).
RN [16]
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;
RA Levy-Adam F., Abboud-Jarroos G., Guerrini M., Beccati D.,
RA Vlodaysky I., Ilan N.;
RT "Identification and characterization of heparin/heparan sulfate
RT binding domains of the endoglycosidase heparanase.";
RL J. Biol. Chem. 280:20457-20466(2005).
RN [17]
RN VARIANT SER-260.
RX PubMed=15334672;
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
RA Shi L.H., Shen H.W., Xu B.;
RT "Heparanase mRNA expression and point mutation in hepatocellular
RT carcinoma.";
RL World J. Gastroenterol. 10:2795-2799(2004).
RN [18]
RN FUNCTION: Endoglycosidase which is a cell surface and
RN extracellular matrix-degrading enzyme. Cleaves heparan sulfate
RN proteoglycans (HSPGs) into heparan sulfate side chains and core
RN proteoglycans. Also implicated in the extravasation of leukocytes
RN and tumor cell lines. Due to its contribution to metastasis and
RN angiogenesis, it is considered to be a potential target for anti-
RN cancer therapies.
RN [19]
RN ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
RN lower extent, by heparin and sulfamin and activated by calcium and
RN magnesium (By similarity).
RN [20]
RN BIOPHYSICO-CHEMICAL PROPERTIES:
RN pH dependence:
RN Optimum pH is 4-6;
RN [21]
RN SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
RN subunits, the proteolytic products.
RN [22]
RN SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
RN [23]
RN Secreted, internalised and transferred to late endosomes/lysosomes

CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium.
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
CC weakly expressed in lymph node, thymus, peripheral blood
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor
CC tissues.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme.
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.
Query Match 75.9%; Score 2126; DB 1; Length 543;
Best Local Similarity 75.9%; Pred. No. 1.8e-150; Indels 0; Gaps 0;
Matches 406; Conservative 50; Mismatches 79;
QY 2 LRPELLLLWMLGRRLALTOGTAPGTAPKVVVDLEFYTKRLFQSPSPSLTIDASLATD 61
DB 9 LPPPLMLLLGLGLPLSPGALPRPAQAQVVDLFFFTQEPHLHLYSPSPSLVTIDANLATD 68
QY 62 PRFLTFLGSPRLALARGLSPAYLRFQGTKTDFLIFDPNKEPTSEERSYWSQDNNDICG 121
DB 69 PRFLILGSPKRLTARGLSPAYLRFQGTKTDFLIFDPKKEPTSEERSYWSQDNNDICG 128
QY 122 SERVSADVLKQLQWEPFOELLILLREOYREFKNSTYSRSSVDMLYSPAKSRDLIFGL 181
DB 129 YGSIPDPVEEKLLEWYQQLLLREHYQKKFNSTYSRSSVDVLYTFANCGLDLIFGL 188
QY 182 NALLRTPDLRWNSNAOLLNLYCSSKGYNTSWELGNEPNSFWKKAQISIDGLQLGEDFE 241
DB 189 NALLRTADLQWNSNAOLLNLYCSSKGYNTSWELGNEPNSFLKADIFINGSQLEDPIQ 248
QY 242 LHKLLQKSAFONAKLYGPDIGQPRGKTVKLLRSLFKAGGEVIDSLTWHYYLNGRVATKE 301
DB 249 LHKLLRKSFTFNKALYGPVQVQPRRTAKMLKSLFKAGGEVIDSVTWHYYLNGRTATRE 308
QY 302 DFLSSDVLDTFILSVQKILKVTKEWTPGKVKWLGESTSAYGGGAPLISNTFAAGFMWLDK 361
DB 309 DFLNDPVDLDTFISVQV 368
QY 362 LGLSAOLGIEVMVRQVFFGAGNHLVDENPEPLDYWLILLFKLVGPKVLMRSRVKGPDR 421
DB 369 LGLSARMGIEVMVRQVFFGAGNHLVDENPDPLDYWLILLFKLVGPKVLMASVQGSKR 428
QY 422 SKLRVYLHCTNVYHPRYREGDLTYLVNLHNVTYKHLKLPMPFSPRPVDKYLKPKFGSGL 481
DB 429 RKLRYVLHCTNTDNPYKEGDLTYLVNLHNVTYKHLKLPMPFSPRPVDKYLKLPFGHL 488
QY 482 LSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKTAACI 536
DB 489 LSKSVQLNGTLKMWDDQTLPLMEKPLRPGSSILGPAFSYGFVIRNAKVAACI 543
RESULT 10
ID HPSE_CHK1
AC Q30YK5
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].

Qy	281	EVIDSLTWHYHYNLGRVATKEDFLSSDVLDTFILSVQKILKVTWKEMTPGKVKWLGETSSA	340
Dd	244	EAVDACTWHYHYNLGDREASLEDFLDPDVLDTLREKIGELVEHVQVSPGKPVWLGETSSA	303
Qy	341	YGGGAPLLSNTFPAAGFMWLDKLGLSAQLGIEVVMQVFPFGAGNYHLVDENFELP----	395
Dd	304	YGGGAAGLSDTFVAGFMWLDKLGLAATLGLGLVMQVFLGAGSYHLMDNDLPLPRSGLL	363
Qy	396	--DYWLSLLFFKKLVGPKVLMGR-VKGPDRS-KLRVYLHCTN-----	432
Dd	364	LQDTWLSLLYKRLVGQEVLTHTTGPAGSERVRLYLHCANKQRCSLLOFLSVRQKREA	423
Qy	433	----VYHPRYREGDITLVYLVNLHNVTKHLKLPMPFSPVDKYLLKPF--GSDGLLSKSV	486
Dd	424	RFLSVLSCSYRSGAATLSMNLKQPARISLPRILSSSTVEAFVLESEQPGEGLSRRAR	483
Qy	487	QLNGQTLKNWDEQTLPALTEKPLPAGSSLSVPAPSYFGFFVIRNAKIAACI	536
Dd	484	KINGRVLMRVODETFPELEGRSLPAAEHLQLPAYSLAFFVFTDAQAAGCV	533
RESULT 12			
HPSE2 HUMAN			
ID	HPSE2 HUMAN	STANDARD;	PRT; 592 AA.
AC	Q5VUH2; Q5VUH4; Q5VUH5; Q5VUH6; Q8WQ01; Q9HB37; Q9HB38; Q9HB39;		
DT	25-OCT-2005, integrated into UniProtKB/Swiss-Prot.		
DT	25-OCT-2005, sequence version 2.		
DT	07-MAR-2006, entry version 16.		
DE	Heparanase-2 (EC 3.2.-.-) (Hpa2).		
GN	Name=HPSE2; Synonyms=HPA2;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;		
OC	Homo.		
OX	NCBI_TaxID=9606;		
ON	[1]		
RP	NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 3 AND 4), TISSUE SPECIFICITY,		
RP	AND SUBCELLULAR LOCATION.		
RC	TISSUE=Heart;		
RX	MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbrc.2000.3586;		
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,		
RA	Hirocock M., Patel S., Barry E., Scubberfield C., Terrett J., Page M.;		
RT	"Cloning and expression profiling of Hpa2, a novel mammalian		
RL	heparanase family member".		
RL	Biochem. Biophys. Res. Commun. 276:1170-1177(2000).		
RN	[2]		
RN	NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).		
RC	TISSUE=Prostate;		
RA	Legoux P., Legoux R., O'Brien D., Salome M.;		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBS databases.		
RN	[3]		
RN	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TVR-579.		
RP	PubMed=15164054; DOI=10.1038/nature02462;		
RA	Deloukas P., Earthworm M.E., Grafham D.V., Rubinfeld M., French L.,		
RA	Loukas C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,		
RA	Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.L.,		
RA	Taylor A., Batties J., Bird C.F., Alincough R., Almeida J.P.,		
RA	Ashwell R.I.S., Ambrose K.D., Babbage A.K., Baguley C.L., Bailey J.,		
RA	Baranjee R., Bates K., Beasley H., Bray-Allen S., Brown A.J.,		
RA	Brown J.Y., Burford D.C., Burdill W., Burton J., Cahill P., Camire D.,		
RA	Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,		
RA	Corby N., Coulson A., Dhani P., Dutta I., Dunn M., Faulkner L.,		
RA	Frankish A., Frankland J.A., Garner P., Garnett J., Gribble S.,		
RA	Griiffiths C., Grocock R., Gusafson E., Hammond S., Harley J.L.,		
RA	Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.J.,		
RA	Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,		
RA	Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,		
RA	Lee H.M., Leongamornlert D.A., Laird G., Lloyd C., Lloyd D.M.,		
RA	Loveland J., Lovell J., McLaren S., McLay K.E., McMurray A.,		
RA	Maashreghi-Mohammadi M., Matthews L., Milne S., Nickerson T.,		
RA	Nguyen M., Overton-Larty E., Palmer S.A., Pearce A.V., Peck A.I.,		
RA	Pelán S., Phillimore B., Porter K., Rice C.M., Roqosin A., Ross M.T.,		

```

RA Sarafidou T., Sehra H.K., Showkeen R., Suke C.D., Smith M.,
RA Standing L., Sycamore N., Tester J., Thorpe A., Torcasso W.,
RA Tracey A., Tromans A., Tsolas J., Wall M., Walsh J., Wang H.,
RA Weinstock K., West A.P., Willey D.L., Whitehead S.L., Wilming L.,
RA Wray P.W., Young L., Chen Y., Lovering R.C., Moschonas N.K.,
RA Siebert R., Fechtke K., Bentley D., Durbin R., Hubbard T.,
RA Doucette-Stamm L., Beck S., Smith D.R., Rogers J.,
RT "The DNA sequence and comparative analysis of human chromosome 10.";
RL FUNCTION: 429:375-381(2004).
CC -!- NUCRION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=HPA2c;
CC IsoId=Q8WWQ2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WWQ2-2; Sequence=VSP_015852, VSP_015853;
CC Name=3; Synonyms=HPA2b;
CC IsoId=Q8WWQ2-3; Sequence=VSP_015851;
CC Name=4; Synonyms=HPA2a;
CC IsoId=Q8WWQ2-4; Sequence=VSP_015850;
CC -!- TISSUE SPECIFICITY: Widely expressed, with the highest expression
CC in brain, mammary gland, prostate, small intestine, testis and
CC uterus. Found both in normal and cancer tissues.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

```

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
Distributed under the Creative Commons Attribution-NoDerivs License

CC	EMBL	AF282885	AAG23421.1	-	mRNA.
CC	EMBL	AF282886	AAG23422.1	-	mRNA.
CC	EMBL	AF282887	AAG23423.1	-	mRNA.
CC	EMBL	AJ299719	CAC82491.1	-	mRNA.
CC	EMBL	AJ299720	CAC82492.1	-	mRNA.
DR	EMBL	AL590036	CAH73137.1	-	Genomic_DNA.
DR	EMBL	AL139243	CAH73137.1	JOINED	Genomic_DNA.
DR	EMBL	AL356220	CAH73137.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAH73137.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAH73137.1	JOINED	Genomic_DNA.
DR	EMBL	AL139243	CAL14146.1	-	Genomic_DNA.
DR	EMBL	AL356220	CAL14146.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAL14146.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAL14146.1	JOINED	Genomic_DNA.
DR	EMBL	AL590036	CAL14146.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAH70448.1	-	Genomic_DNA.
DR	EMBL	AL139243	CAH70448.1	JOINED	Genomic_DNA.
DR	EMBL	AL356220	CAH70448.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAH70448.1	JOINED	Genomic_DNA.
DR	EMBL	AL590036	CAH70448.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAL16472.1	-	Genomic_DNA.
DR	EMBL	AL139243	CAL16472.1	JOINED	Genomic_DNA.
DR	EMBL	AL356220	CAL16472.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAL16472.1	JOINED	Genomic_DNA.
DR	EMBL	AL590036	CAL17160.1	-	Genomic_DNA.
DR	EMBL	AL139243	CAL17160.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAL17160.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAL17160.1	JOINED	Genomic_DNA.
DR	EMBL	AL590036	CAL17160.1	JOINED	Genomic_DNA.
DR	EMBL	AL590036	CAH73139.1	-	Genomic_DNA.
DR	EMBL	AL139243	CAH73139.1	JOINED	Genomic_DNA.
DR	EMBL	AL356220	CAH73139.1	JOINED	Genomic_DNA.
DR	EMBL	AL356268	CAH73139.1	JOINED	Genomic_DNA.
DR	EMBL	AL445251	CAH73139.1	JOINED	Genomic_DNA.
DR	EMBL	AL356220	CAL17162.1	-	Genomic_DNA.
DR	EMBL	AL356268	CAL17162.1	JOINED	Genomic_DNA.
DR	EMBL	AL139243	CAL17162.1	JOINED	Genomic_DNA.

12 GRLRALTOGTPAGTAPT KDVV DLEFYTKRLFQSVSPFLSITIDASLATDPRFLTFLGSP 71

[illegible]

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
NCBI_TaxID=99883;

[1] NUCLEOTIDE SEQUENCE.

RX PubMed=1549614; DOI=10.1038/nature03025;

RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozout-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dessat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Guzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Reest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype." ;
NT Nature 431:946-957(2004).

[2]

RN NUCLEOTIDE SEQUENCE.

RP GenomeScope; Whitehead Institute Centre for Genome Research;
RG Submitted (FEB-2004) to the ENBL/GenBank/DBSJ databases.
RL -! CAUTION: The sequence shown here is derived from an
CC ENBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.

CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC
DR NON TER 597 597
FT FT
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query March 37.4%; Score 1048; DB 2; Length 597;
Best Local Similarity 40.9%; Pred. No. 1.4e-69;
Matches 236; Conservative 101; Mismatches 196; Indels 44; Gaps 13

Qy 1 MLRPLLLWLWGRLALTOGTAGTAP---TKVDVDFYTKRLFQSVPSPFLSITIDAS 57
Db :|::||:
Qy 22 LLVPLVLSSPVSSSTVQRPAVGKRCFGVFERTILLDVNTSPIRLVNDNFSLQLDPS 81
Db :|::||:
Qy 58 LATDPRELTFIGSPRLRALAGLSPAYLRFCGKTDTFLIPDNKEPTSEER----SYWOS 113
Db :|::||:
Qy 82 IIKD-GWLDDFLSGRLVTLAGLSPAFLRFPGCKRTDLQF-TNQKNLAKFRRGGPDYLK 139
Db :|::||:
Qy 114 QDNNDICGSE-----RVSNADVLRLQMWPFO-EALLIREOYOREFNST---- 157
Db :|::||:
Qy 140 NYEDDIIRSDIALDKQCKGLASHPHMDMLEIQREKAASTQLVLLKQELSNITYSNITLTGI 199
Db :|::||:
Qy 158 YRSR-----SVDMLYSFACSRDLIFLGNALRLTPDLRWNSSNAOLLNYCCKGYNIWS 213
Db :|::||:
Qy 200 FSHSRARSCLKYNFADCAGLHILGLINALHRNPDHSMWTNSTLSLLKYSGAKYNIWS 259
Db :|::||:
Qy 214 ELGNPNFSFWKQAIGSIDGLGBDFVELHLKLQK-SAFONAKLYGPDIGQPRGTVKLL 272
Db :|::||:
Qy 260 ELGNFPNAYRSMVGHAVNSQLAQDYTKRTLTLQSVRYYSRAQLYCPNAGRPRKNALLL 319
Db :|::||:
Qy 273 RSFLKAGEVIDSLTHWHYYLNGVATKDEFLSSDVLDTFILSVOKILKVTEMTPGKKV 332
Db :|::||:
Qy 320 DEFWKTVGVTDVAQTWQHYGMGRICKVEDFKTRLLDTRLTQLSKVTKVNWHTPGKKV 379
Db :|::||:
Qy 333 WLGETSSAYCGGAPILLSNTFAAGFMWLKDGLSLGAOLGIEVMNR----QVFFGAGNYHLVD 388
Db :|::||:
Qy 380 WLGGGLGPAWTGMSNLSDTFAAAGFLVNVTGLMAAMQGIDVVLRROQAOEHNTKQSVLF 439
Db :|::||:
Qy 389 ENFEP-LPDYWLSLLFKLVGPKULMSRVKG-----PD-----RSKLRVYLHCNVHPRY 438
Db :|::||:
Qy 440 QMFVPSFDYWFSLVFRKLVGPKVALAVRAGLQRKPQGRVIRDKLRIYAHCTSYSNHY 499
Db :|::||:
Qy 439 REGDLITLVNLHNVTXHLKLPMPFMRSPVKYLLKPFSGDGLLSKSVOLNGTLKWVDE 498
Db :|::||:

```
Db 500 VRSIIYIIINLHRSRKKIKLACTLRNNIVHQLQPYGAGDLRAKHVQLNGEKLIMADN 559
Qy 499 QTLPALTEKPLPAGSSLSVPFSGVFFVIRNAKTAAC 535
Db 560 ETPPELKPKTLRAGRTIAMPMTTIGTFVVIKNINAYAC 596

RESULT 15
Q4TGC8 TETNG
ID Q4TGC8 TETNG PRELIMINARY; PRT; 255 AA.
AC Q4TGC8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00001168001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Luffalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAE01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

Query Match 25.9%; Score 725.5; DB 2; Length 255;
Best Local Similarity 54.7%; Pred. No. 6.5e-46;
Matches 139; Conservative 43; Mismatches 55; Indels 17; Gaps 3;

Qy 159 SRSSVDMLYSFACSRDLIFGLNALLRTPDLRWNSSNAQLLNYCCKGNYISWEIGNE 218
Db 1 SETTVQLHAFACSGGLDLVFLGNALLRTADNRWSSNARSLRLRYCARRYHMSWEIGNE 60
Qy 219 PMSFWKKAQISIDGLQGEDFVELKHLQKSAF-ONAKLYGPDIGQPRGKTVKLLRSFLK 277
Db 61 PMSYKKAGRLDGRQLGEDFTVLRKILRESRFRDAGLFGPDVGQPRDRHIDILSGFLQ 120
Qy 278 AGGEVIDSLTWHYHLYNGRVATKEDFLSSDVLDTFLISVQKILKVKEMTPGKKVWLGET 337
Db 121 SGAEAVDACTWHYHLYLDGREASLEDFLDPDVLDTLRKICEVLEEVHQSFGKPVWLGET 180
```

```
Qy 338 SSAYGGG-----APLLSNTFAAG-PMWLDKLGLSAQLGIEVVMRQVFFGA 381
Db 181 SSATGAEPGRGCRTHSSQDSCFAPRRSDOAPLGTFRWLDKLGAAATLGLELYMRQVLGA 240
Qy 382 GNYHLVDENPEPLP 395
Db 241 GSYHLMDDNLDPLP 254
```

Search completed: June 5, 2006, 12:20:15
Job time : 134.991 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 132.743 Seconds
(without alignments)
3728.138 Million cell updates/sec

Title: US-10-645-659A-2
Perfect score: 2797
Sequence: 1 MLRLLLLLWGLGALQA.....LPAFSYGFFVIRNAKIAACI 535

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 7.2.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2775	99.2	535	1	HPSE_MOUSE
2	2601.5	93.0	536	1	HPSE_RAT
3	2392.5	85.5	574	2	Q333x8 GRODE
4	2392.5	85.5	574	2	Q333x9 spalax gali
5	2387.5	85.4	574	2	Q333x7 GRODE
6	2379.5	85.1	574	2	Q333x6 spalax juda
7	2274.5	81.3	558	2	Q333x5 SPAJD
8	2184	78.1	545	1	HPSE_BOVIN
9	2149	76.8	543	1	HPSE_HUMAN
10	1605	57.4	523	1	HPSE_CHICK
11	1289	46.1	533	2	Q4SF6_TETNG
12	1169.5	41.8	592	1	HPSE2_HUMAN
13	1169.5	41.8	592	2	Q2MIH9 HUMAN
14	1042.5	37.3	597	2	Q4T880 TETNG
15	724.5	25.9	255	2	Q4TG8_TETNG
16	671	24.0	515	2	Q8T108_BOMMO
17	415	14.8	559	2	Q89F99_BRAJA
18	391	14.0	543	1	HPSE1_ARATH
19	372	13.3	527	2	Q9LRC8_SCUBA
20	362	12.9	526	2	Q5SNA6_ORYSA
21	361.5	12.9	541	2	Q69I16_ORYSA
22	357.5	12.8	539	2	Q2QN56_ORYSA
23	355.5	12.7	536	1	HPSE3_ARATH
24	346.5	12.4	529	2	Q6ZJE2_ORYSA
25	346.5	12.4	537	2	Q70YJ3_HORVU
26	336	12.0	401	2	Q30324_ARATH
27	335.5	12.0	516	2	Q447R5_SOLUS
28	310.5	11.1	539	1	HPSE2_ARATH
29	309.5	11.1	537	2	Q43S03_SOLUS
30	275.5	9.8	506	2	Q37Q70_SPHAR
31	242	8.7	382	2	Q3E8P7_ARATH

32	153	5.5	510	2	Q2U0T3 ASPOR	Q2u0t3 aspergillus
33	151	5.4	1128	2	OSTT65 ANOGA	OSTt65 anopheles g
34	138	4.9	559	2	Q7SP80_NEUCR	Q7sfb0 neurospora
35	135.5	4.8	935	2	Q9VE79_DROME	Q9ve79 drosophila
36	133	4.8	670	2	Q3JTG0_BURP1	Q3jtg0 burkholderi
37	132	4.7	463	2	Q63T97_BURPS	Q63t97 burkholderi
38	130	4.6	1053	2	P77865_FIBSU	P77865 fibrobacter
39	129	4.6	1053	2	P71329_FIBSU	P71329 fibrobacter
40	124.5	4.5	1303	2	Q2ZH58_CALSA	Q2zh58 caldicellul
41	122.5	4.4	628	2	Q2ZEV8_CALSA	Q2zev8 caldicellul
42	118.5	4.2	536	2	Q2UDS9_ASPOR	Q2uds9 aspergillus
43	118.5	4.2	721	2	Q3XV07_9PROT	Q3xv07 magnetococc
44	118	4.2	2638	2	Q551W7_CRYNE	Q551w7 cryptococcu
45	118	4.2	2638	2	Q5KCT5_CRYNE	Q5ket5 cryptococcu

ALIGNMENTS

RESULT 1

ID HPSE_MOUSE STANDARD; PRT; 535 AA.
AC Q6YGZ1, Q8K3K3;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 13.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).
GN Name=Hpse; Synonyms=Hpa;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=SJL/J; TISSUE=Spleen; DOI=10.1038/10525;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1074/jbc.M300925200;
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RL Nat. Med. 5:803-809(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,
RP GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND
RP SUBUNITS.
RC STRAIN=FVB; TISSUE=Embryo;
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
RT "Cloning, expression, and purification of mouse heparanase";
RT Protein Expr. Purif. 26:425-431(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,
RA Li J.-P.;
RT "Processing of macromolecular heparin by heparanase";
RT J. Biol. Chem. 278:35152-35158(2003).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RX STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
RC PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi S., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels B.P., de Bono B., Della Gatta G.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,

DT	11-OCT-2005, integrated into UniProtKB/Swiss-Prot.	FT	ACT_SITE	218	218	Proton donor (Potential).
DT	05-JUL-2004, sequence version 1.	FT	ACT_SITE	336	336	Nucleophile (Potential).
DT	07-MAR-2006, entry version 11.	FT	CARBOHYD	155	155	N-linked (GlcNAc. . .) (By similarity).
DE	Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:	FT	CARBOHYD	193	193	N-linked (GlcNAc. . .) (By similarity).
DE	Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].	FT	CARBOHYD	210	210	N-linked (GlcNAc. . .) (By similarity).
GN	Name=Hspe; Synonyms=Hep;	FT	CARBOHYD	452	452	N-linked (GlcNAc. . .) (By similarity).
GN	Rattus norvegicus (Rat);	FT	CONFLICT	15	15	G -> R (in Ref. 2).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	CONFLICT	227	227	H -> Q (in Ref. 2).
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	FT	CONFLICT	350	350	D -> N (in Ref. 2).
OC	Muroidea; Muridae; Murinae; Rattus.	SQ	SEQUENCE	536 AA; 60480 MW; C434E04CF536EA4D CRC64;		
OX	NCBI_TaxID=101116;					
OX	NCBI_TaxID=101116;					
RP	NUCLEOTIDE SEQUENCE [MRNA].					
RC	TISSUE=Placenta;					
RC	MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;					
RX	Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,					
RA	Parish C.R.;					
RA	"Cloning of mammalian heparanase, an important enzyme in tumor					
RT	invasion and metastasis.";					
RT	Nat. Med. 5:803-809(1999).					
RL	[2]					
RP	NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.					
RP	MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;					
RA	Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;					
RX	"Characterization of heparanase from a rat parathyroid cell line.";					
RT	J. Biol. Chem. 277:32459-32465(2002).					
CC	-!- FUNCTION: Endoglycosidase which is a cell surface and					
CC	extracellular matrix-degrading enzyme. Cleaves heparan sulfate					
CC	proteoglycans (HSPGs) into heparan sulfate side chains and core					
CC	proteoglycans. Also implicated in the extravasation of leukocytes					
CC	and tumor cell lines. Contributes to metastasis and angiogenesis					
CC	(By similarity).					
CC	-!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower					
CC	extent, by heparin and sulfamin (By similarity). Activated by					
CC	calcium and magnesium. Inhibited by EDTA.					
CC	-!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa					
CC	subunits, the proteolytic products (By similarity).					
CC	-!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.					
CC	Secreted, internalised and transferred to late endosomes/lysosomes					
CC	as a proheparanase. In lysosomes, it is processed into the active					
CC	form, the heparanase. The uptake or internalisation of					
CC	proheparanase is mediated by HSPGs. Heparin appears to be a					
CC	competitor and retain proheparanase in the extracellular medium					
CC	(By similarity).					
CC	-!- PTM: Proteolytically processed. The cleavage of the 65 kDa form					
CC	leads to the generation of a linker peptide, 8 kDa and 50 kDa					
CC	product. The active form, the 8/50 kDa heterodimer, is resistant					
CC	to degradation. Complete removal of the linker peptide appears to					
CC	be a prerequisite to the complete activation of the enzyme (By					
CC	similarity).					
CC	-!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears					
CC	to be essential for its solubility (By similarity).					
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.					
CC	-----					
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms					
CC	Distributed under the Creative Commons Attribution-NoDerivs License					
CC	-----					
DR	EMBL; AF359508; AAQ15189.1; -; mRNA.					
DR	EMBL; AF184967; AAF04563.1; -; mRNA.					
DR	RGD; 61969; Hspe.					
DR	InterPro; IPR005199; Glyco_hydro_79_N.					
DR	Pfam; PF03662; Glyco_hydro_79n; 1.					
KW	Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;					
FT	SIGNAL.					
FT	CHAIN 1 28 By similarity.					
FT	CHAIN 29 102 Heparanase 8 kDa subunit.					
FT	PROPEP 103 150 /Frid-PRO_0000042266.					
FT	CHAIN 151 536 /Frid-PRO_0000042267.					
FT	CHAIN 151 536 Heparanase 50 kDa subunit.					
FT	REGION 151 155 /Frid-PRO_0000042268.					
FT	REGION 263 273 Heparin/HS-binding (By similarity).					


```

RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48EBEFC7D0BCB34 CRC64;

Query Match      85.5%; Score 2392.5; DB 2; Length 574;
Best Local Similarity 84.9%; Pred. No. 5.7e-172;
Matches 455; Conservative 35; Mismatches 43; Indels 3; Gaps 2;

QY 1 MLRL-LLLWLWGLGALAGAPAGTAGTDDVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
DB 41 MLRLSLLLWLWGLSPLVQCILA--AQAEVVELEFSTQRPPLHVSFSLITIDANLAT 98
QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLFDPPDKEPTSEERSYKWSQVNHDI 119
DB 99 DPRFLTFLGSPKLALARGLSPAYLRFGGTKTDFLFDPPKEPSHEERSYKWSQVNHDI 158
QY 120 RSEPVSAAVLRKLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYSPAKCSGLDLIFG 179
DB 159 RSGAIPAVVVRRLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYTFARCSGLDLIFG 218
QY 180 LNALRTDPLRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
DB 219 LNALRTADFRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLISIDGLQGEDYI 278
QY 240 ELHKLRLQSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLLNGRIATK 299
DB 279 ELHKLRLKSTLKNVLYGPDVGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLLNGRIATK 338
QY 300 EDFLSSDALDFTILSVOKILKVTKEITPGKKVWLGTSAYGGGAPLLSNTFAAGFMWLD 359
DB 339 EDFLSPDVLDTFILSVOKILQVVEETRPKKVWLGTSAYGGGAPLLSNTFAAGFMWLD 398
QY 360 KLGLSAQMGIENVMRQVFFGAGNYHLVDENFEPDPYWLSSLFKKLGVPRVLLSRVKGPD 419
DB 399 KLGLSQMGIENVMRQVFFGAGNYHLVDKNFEPDPYWLSSLFKKLGVSKVLMARVKGPD 458
QY 420 RSKLRVYLHCTNVYHPRYQEGDLTLVNLNHNVTKHLKVPDPFRKPKVDYTLKPSGPDG 479
DB 459 RSKLRVYLHCTNVHPRYQEGDLTLVNLNHNVTKHLKVPDPFRKPKVDYTLKPLGPG 518
QY 480 LLSKSVQLNGQILKMVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIACI 535
DB 519 LLSKSVQLNGQALKMVDQTLPALTEKPLRPGSSGLPFAFSYGFFVIRNAKVAACL 574

RESULT 4
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
```

```

RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      85.5%; Score 2392.5; DB 2; Length 574;
Best Local Similarity 84.9%; Pred. No. 5.7e-172;
Matches 455; Conservative 35; Mismatches 43; Indels 3; Gaps 2;

QY 1 MLRL-LLLWLWGLGALAGAPAGTAGTDDVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
DB 41 MLRLSLLLWLWGLSPLVQCILA--AQAEVVELEFSTQRPPLHVSFSLITIDANLAT 98
QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLFDPPDKEPTSEERSYKWSQVNHDI 119
DB 99 DPRFLTFLGSPKLALARGLSPAYLRFGGTKTDFLFDPPKEPSHEERSYKWSQVNHDI 158
QY 120 RSEPVSAAVLRKLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYSPAKCSGLDLIFG 179
DB 159 RSGAIPAVVVRRLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYTFARCSGLDLIFG 218
QY 180 LNALRTDPLRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
DB 219 LNALRTADFRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLISIDGLQGEDYI 278
QY 240 ELHKLRLQSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLLNGRIATK 299
DB 279 ELHKLRLKSTLKNVLYGPDVGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLLNGRIATK 338
QY 300 EDFLSSDALDFTILSVOKILKVTKEITPGKKVWLGTSAYGGGAPLLSNTFAAGFMWLD 359
DB 339 EDFLSPDVLDTFILSVOKILQVVEETRPKKVWLGTSAYGGGAPLLSNTFAAGFMWLD 398
QY 360 KLGLSAQMGIENVMRQVFFGAGNYHLVDENFEPDPYWLSSLFKKLGVPRVLLSRVKGPD 419
DB 399 KLGLSQMGIENVMRQVFFGAGNYHLVDKNFEPDPYWLSSLFKKLGVSKVLMARVKGPD 458
QY 420 RSKLRVYLHCTNVYHPRYQEGDLTLVNLNHNVTKHLKVPDPFRKPKVDYTLKPSGPDG 479
DB 459 RSKLRVYLHCTNVHPRYQEGDLTLVNLNHNVTKHLKVPDPFRKPKVDYTLKPLGPG 518
QY 480 LLSKSVQLNGQILKMVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIACI 535
DB 519 LLSKSVQLNGQALKMVDQTLPALTEKPLRPGSSGLPFAFSYGFFVIRNAKVAACL 574

RESULT 5
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN
```

```

RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RL cloning and identification of a novel splice variant.";
Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCBADD99DE CRC64;

Query Match      85.4%; Score 2387.5; DB 2; Length 574;
Best Local Similarity 84.9%; Pred. No. 1.4e-171;
Matches 455; Conservative 34; Mismatches 44; Indels 3; Gaps 2;

QY 1 MLRL-LLLWLGPGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 MLRLSLLLWLGPLSPVQCILA--AQAEVVLEFSTQRLPLHLVSPFLSITIDANLAT 98

QY 60 DPRFLTFLGSPRLALARGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDC 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 DPRFLTFLGSPKALARGLSPAYLRFGGTKTDFLI FDPKPEPSHEERSYWKSVQVNHDC 158

QY 120 RSEPVSAVLKQVLEWPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFG 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 RSGAIPAVVVRRLQVLEWPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFG 218

QY 180 LNALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 LNALLRTADFRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDYI 278

QY 240 ELHKLQSFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 ELKLLRKSTLKNVLYGPDVGQPRGTVKLLRSFLKAGGEVIDSVTWHHYLLNGRIATK 338

QY 300 EDFLSSDALDTFILSVQKILKVTKETITPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 KLGLSAQMGIEMVVRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGSKVLMARVKGPD 458

QY 420 RSKLRVYLHCTNVHPRYQEGDLTLVYLNHNVTGHLKVPPLPRKVPDVTYLLKPSGPDG 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 RSKLRVYLHCTNINHPRYQEGDLTLVYALNLYNTGHLKLPYQLFNKPVDKYLVIPLGPGG 518

QY 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 LLSKSVQLNGQALKWDDQTLPALTEKPLRPGSSLSGLPAFSYGFFVIRNAKVAACL 574

RESULT 6
Q333X6 SPAJD
ID Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene

```

```

RT cloning and identification of a novel splice variant.";
Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RL cloning and identification of a novel splice variant.";
Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3ABEB13F07451684 CRC64;

Query Match      85.1%; Score 2379.5; DB 2; Length 574;
Best Local Similarity 84.7%; Pred. No. 5.5e-171;
Matches 454; Conservative 34; Mismatches 45; Indels 3; Gaps 2;

QY 1 MLRL-LLLWLGPGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 MLRLSLLLWLGPLSPVQCILA--AQAEVVLEFSTQRLPLHLVSPFLSITIDANLAT 98

QY 60 DPRFLTFLGSPRLALARGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDC 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 DPRFLTFLGSPKALARGLSPAYLRFGGTKTDFLI FDPKPEPSHEERSYWKSVQVNHDC 158

QY 120 RSEPVSAVLKQVLEWPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFG 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 RSGAIPAVVVRRLQVLEWPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFG 218

QY 180 LNALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 LNALLRTADFRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDYI 278

QY 240 ELHKLQSFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 ELKLLRKSTLKNVLYGPDVGQPRGTVKLLRSFLKAGGEVIDSVTWHHYLLNGRIATK 338

QY 300 EDFLSSDALDTFILSVQKILKVTKETITPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 KLGLSAQMGIEMVVRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGSKVLMARVKGPD 458

QY 420 RSKLRVYLHCTNVHPRYQEGDLTLVYLNHNVTGHLKVPPLPRKVPDVTYLLKPSGPDG 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 RSKLRVYLHCTNINHPRYQEGDLTLVYALNLYNTGHLKLPYQLFNKPVDKYLVIPLGPGG 518

QY 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 LLSKSVQLNGQALKWDDQTLPALTEKPLRPGSSLSGLPAFSYGFFVIRNAKVAACL 574

RESULT 7
Q333X5 SPAJD
ID Q333X5 SPAJD PRELIMINARY; PRT; 558 AA.
AC Q333X5
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```


Db 74 FTFLGSKLRTLARGAPAYLRFGNGKDFLLFPDKPEAFERSYWLQSNQDICKSGS 133
Qy 124 VSAVLKQVQPPQELLRLRQYKFKPNSTYRSRSDMLYSFAKCSGLDLIFGNAL 183
Db 134 IPSDVBEKLRLWPFQBVLLRQYKFKTNSTYRSRSDMLYTFASCGLNLIIFGNAL 193
Qy 184 LRTPTDLRWSSNAQLLDYCSSKYNINLWELGNEPNSFWKKAHLIDLOLGEDFVFLHK 243
Db 194 LRTTDMHDSNAQLLDYCSSKYNINLWELGNEPNSFORKAGIFINGROLGEDFIFBRK 253
Qy 244 LLORSFONAKLYGPDIGQPRGTQVLLRSFLKAGGEVIDSLTWHYYLNGRIATKEDFL 303
Db 254 LLKSFAFKNAKLYGPDIGQPRNTVWLKSLFKAGGEVIDSVTWHYYVNGRATKEDFL 313
Qy 304 SSDALDTFILSVQKILKVTKEITPGKVKWLGTSAYGGGAPLLNSNTFAAGFWMLDKGL 363
Db 314 NPDILOTFISSVQKTRIVEKIRPLKVKWLGTSAYGGGAPLLNSNTFAAGFWMLDKGL 373
Qy 364 SAQMGIEVMRQVFFGAGNYHLVDENPEPLPDYWLSSLFKKLVGPRVLLSRVKGPDRLSKL 423
Db 374 SARMGIEVMRQVLFAGNYHLVDGNFPEPLPDYWLSSLFKKLVGNKVMASVKGPDRLSKF 433
Qy 424 RYVLHCTNVYHPRYQSGDLTYVLNHLNVTKHLKVPPELPRKPVDTYLLKPSGPDGLLSK 483
Db 434 RYVLHCTNTKHPRYKGGDLTYALNHLNVTKHLKVPPELPRKPVDTYLLKPSGPDGLLSK 493
Qy 484 SVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSGYGFVFNRAKIAACI 535
Db 494 SVQLNGQILKMWDEQTLPALTEKPLPAGSLSGMPFSGYGFVFNRAKIAACI 545

RESULT 9

HPSE HUMAN
ID HPSE HUMAN STANDARD; PRT; 543 AA.
AC Q9Y2E1; Q53G5; Q9UL39;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1999, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Heparanase precursor [EC 3.2.-.-] (Heparanase-1) (Hpal) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
DE subunit].
GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPR1, HPSE1, HSE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
RA Kusile P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and functional expression of a human heparanase gene.";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.
RC TISSUE=Embryonic fibroblast;
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and expression.";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
RX PubMed=10395325; DOI=10.1038/10518;
RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Pecker I.;
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";

Nat. Med. 5:793-802(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
RN [5]
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular damage.";
RL Glycobiology 10:467-475(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
RX PubMed=11547900; DOI=10.1023/A:1011375624902;
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O., Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
RT "Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis.";
RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
RN [7]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163, SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
RC TISSUE=Placenta;
RX PubMed=12713442; DOI=10.1042/BJ20030318;
RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R., Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S., Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M., Parekh R., Stubberfield C.;
RT "Biochemical characterization of the active heterodimer form of human heparanase (Hpal) protein expressed in insect cells.";
RL Biochem. J. 373:423-435(2003).
RN [8]
RP NUCLEOTIDE SEQUENCE [MRNA].
RA Pinhal M.A., Semedo P.;
RT "Cloned heparanase from MCF-7 cells.";
RN Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
RN [9]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Small intestine;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
RN [10]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Beak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP [11]
RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
RX PubMed=11123890; DOI=10.1021/bi002080p;
RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
RA Gready J.E., Farish C.R.;
RT "Identification of active-site residues of the pro-metastatic
RT endoglycosidase heparanase.";
RL Biochemistry 39:15659-15667(2000).
RN [12]
RN N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
RP ASN-217; ASN-238 AND ASN-459.
RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
RA Simizu S., Ishida K., Wierzbza M.K., Osada H.;
RT "Secretion of heparanase protein is regulated by glycosylation in
RT human tumor cell lines.";
RL J. Biol. Chem. 279:44084-44092(2004).
RN [14]
RN SUBCELLULAR LOCATION.
RP PubMed=15292202; DOI=10.1074/jbc.M402131200;
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodaysky I., Ilan N.;
RT "Heparanase uptake is mediated by cell membrane heparan sulfate
RT proteoglycans.";
RL J. Biol. Chem. 279:44084-44092(2004).
RN [14]
RN BIOPHYSICO-CHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
RA Cohen E., Atzmon R., Vlodaysky I., Ilan N.;
RT "Heparanase processing by lysosomal/endosomal protein preparation.";
RL FEBS Lett. 579:2334-2338(2005).
RN [15]
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
RA Abboud-Jarroos G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
RA Elgavish S., Peretz T., Vlodaysky I.;
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
RT human proheparanase.";
RL J. Biol. Chem. 280:13568-13575(2005).
RN [16]
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;
RA Levy-Adam F., Abboud-Jarroos G., Guerrini M., Beccati D.,
RA Vlodaysky I., Ilan N.;
RT "Identification and characterization of heparin/heparan sulfate
RT binding domains of the endoglycosidase heparanase.";
RL J. Biol. Chem. 280:20457-20466(2005).
RN [17]
RN VARIANT SER-260.
RX PubMed=15334672;
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
RA Shi L.H., Shen H.W., Xu B.;
RT "Heparanase mRNA expression and point mutation in hepatocellular
RT carcinoma.";
RL World J. Gastroenterol. 10:2795-2799(2004).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -!- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
CC lower extent, by heparin and sulfamin and activated by calcium and
CC magnesium (By similarity).
CC -!- BIOPHYSICO-CHEMICAL PROPERTIES:
CC pH dependence:
CC Optimum pH is 4-6;
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products.
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes

CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium.
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
CC weakly expressed in lymph node, thymus, peripheral blood
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor
CC tissues.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme.
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.
Query Match 76.8%; Score 2149; DB 1; Length 543;
Best Local Similarity 76.6%; Pred. No. 1.4e-153;
Matches 407; Conservative 50; Mismatches 74; Indels 0; Gaps 0;
QY 5 LLLWLMGPGALGAQAGAPAGTAPTDVVDLEFYTQKPLRSVSPSLSTITDASLATDPRFL 64
DB 13 LMLLLGLPLGSPGALPRPAQADVVLDFFQEPHLHLYSPSLSTITDANLATDPRFL 72
QY 65 TFIGSPRLRALARGSPAYLRFPGTKTDFLI FPDPKPTSEERSYKQSNVNDICRSEPV 124
DB 73 ILLGSPKRLTARGSPAYLRFPGTKTDFLI FDPKKESTFEERSYQSNQDICKYGI 132
QY 125 SAAVLRKLVQVWPFOELLRLLEQYQKFEKNSYSSRVSDMLYSFAKSGLLDIFGLNALL 184
DB 133 PPVVEEKLRLWEPYQEQLLREHYQKKFNSTYSRSSVDVLYTTFANCGLDLIFGLNALL 192
QY 185 RTPDLRWNSNAQLLLDYCSSKGYNISWELGNENPNSFWKKAHLIDGLQLGEDFVELHKL 244
DB 193 RTADLQWNSNAQLLLDYCSSKGYNISWELGNENPNSFKKADI FINGSQLGEDFIQLHKL 252
QY 245 LQSAFQNAKLYGPDIGQPRKTVKLLRSFLKAGEVIDSLTWHYLYNGRIATKEDFLS 304
DB 253 LRKSTFNKAKLYGPDVGQPRRKTA KMLKSLKAGEVIDSVTWHYLYNGRTATREDFLN 312
QY 305 SDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLS 364
DB 313 PDVLDIFISSVQKVFQVVESTTRPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLS 372
QY 365 AOMGIEVWQVFFGAGNYHLVDENFPLPDYWLISLFLKLVGPRVLLSRVKGPDRSKLR 424
DB 373 ARMGIEVWQVFFGAGNYHLVDENFPLPDYWLISLFLKLVGTVKVLMAVQGSKRKL 432
QY 425 VYLHCTNVVHPRYQEGDGLTYLVNLHNVTKHLKVPPLFRKPDVTYLLKPSGPDGLLSKS 484
DB 433 VYLHCTNDNPRYKEGDGLTYLVNLHNVTKYLRPLYPFNSKQVDKYLRLPLGPHGLLSKS 492
QY 485 VOLNGQLKQWDEQTLPTALTEKPLPAGSALSPLAFSYGFFVIRNAKTAACI 535
DB 493 VOLNGLTLKMWDDQTLPLMEKPLRPGSSGLPLAFSYGFFVIRNAKVAACI 543
RESULT 10
ID HPSE_CHICK STANDARD; PRT; 523 AA.
AC Q30YK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].

XX	MedLine=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;	
RA	Goldshmidt O., Zcharia E., Alingorn H., Guatta-Ranghri Z., Atzmon R.,	
RA	Michal I., Fecker I., Mitrani E., Vlodavsky I.,	
RT	"Expression pattern and secretion of human and chicken heparanase are	
RT	determined by their signal peptide sequence.";	
RL	J. Biol. Chem. 276:29178-29187 (2001).	
CC	-1- FUNCTION: Endoglycosidase which is a cell surface and	
CC	extracellular matrix-degrading enzyme. Cleaves heparan sulfate	
CC	proteoglycans (HSPGs) into heparan sulfate side chains and core	
CC	proteoglycans (By similarity).	
CC	-1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted	
CC	(By similarity).	
CC	-1- PTM: N-glycosylated (By similarity).	
CC	-1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.	
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms	
CC	Distributed under the Creative Commons Attribution-NoDerivs License	
CC	-----	
DR	EMBL; AY037007; AAX82648.1; -; mRNA.	
DR	Ensembl; ENSGALG0000011203; Gallus gallus.	
DR	InterPro; IPR005199; Glyco_hydro.79_N.	
DR	Pfam; PF03662; Glyco_hydro.79n; 1.	
KW	Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.	
FT	SIGNAL 1 18 Potential.	
FT	CHAIN 19 523 Heparanase.	
FT	REGION 137 141 /FTID=PRO_0000042259.	
FT	REGION 250 260 Heparin/HS-binding (By similarity).	
FT	ACT_SITE 204 204 Heparin/HS-binding (By similarity).	
FT	ACT_SITE 323 323 Proton donor (Potential).	
FT	ACT_SITE 323 323 Nucleophile (Potential).	
FT	CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).	
FT	CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).	
FT	CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).	
FT	CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).	
SQ	SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;	
Query Match 57.4%; Score 1605; DB 1; Length 523;		
Best Local Similarity 58.8%; Pred. No. 1.9e-112;		
Matches 315; Conservative 81; Mismatches 126; Indels 14; Gaps 4		
QY	1 MLRLILLWGLGALGAQAAGAGTPTDDVDVLEFYTKRPLRSVSPSLFTTIDASLATD 60	
DB	1 MLVLLLLLVL-----LNAVPRRTA-----ELQGLREPAGVSPFLSTLTDASLARD 48	
QY	61 PRFLTFGLSPRLRALARGLSPAYLRFGGKTDFLIFDPDKPTSEERSYKWSQVNHDIR 120	
DB	49 PRFVALLRHPKHLTLASGLSPGFLRFGGTSDFLIFENKDSWEEKVLSFEQAK-DVCE 107	
QY	121 SEPVSAAVLARKLOVEWFPQELLRLREQYQKFKNSTYSRRSSVDMLYSPAKSGLDLIFGL 180	
DB	108 AWPFAVVPKLLLTQWPLQEKLLAEHSKKHKNTTITRSTLDILHTPASSSGFLRVFGL 167	
QY	181 NALLRTFDLRWSSNAQLLLDYCSSKGYNISEWELGNEPNSFWKKAHILLDGLQGEDFVE 240	
DB	168 NALLRRAGLQWSSNAQLLLDYCSSKGYNISEWELGNEPNSFRKSGICIDGFLQGRDFVH 227	
QY	241 LHKLL-QRSAFONAKLVGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLYNGRIATK 299	
DB	228 LRQLLSQHPLYRHAELVGLDVQGRPKHTQHLRSPMKSGGKAIDSVTWHHYVNGRSATR 287	
QY	300 EDFLSSDALDITFILSVQKILKVTKXETIPGKKVWLGETSSAYGGGAPLLISNTFAAGFMWLD 359	
DB	288 EDFLSPEVLDSFATAIHDLVGI VEATVP GKVKWLGETSGAYGGGAPQLSNTVYVGFWMWLD 347	
QY	360 KLGLSAQMGIEVVMRQVFFGAGNTHLVDENTEPLPDYMLSLFLFKLVGRVLLSRVKGPD 419	
DB	348 KLGLAARRGIDVVMRQVSFGAGSYHLVDAGFKPLPDYMLSLLYKRLVGRVLLQASVEQAD 407	
QY	420 RSKLRVYLHCTNVVHPRYOEGDGLTYLVNLHNVTKHLKVPPELPFRKPVDTYLLKSGPDG 479	
DB	408 ARRPVVYLHCTNPRHPKTRGEGDVTYFALNLSNVTQSLOI PQKLWSEKSDVQYLLHPGKDS 467	
QY	480 LLKSKSVQLNGOILKMWDEQTLPALTEKPELPAGSALSLPAFSYGFVIRNAKTAACI 535	

Db 468 ILSREVQLNGRLQLQWDDDELTPALHWAALAPGSLTGLPAFSYGFYVIRNAKATACI 523

RESULT 11

Q4SYF6_TETNG PRELIMINARY; PRT; 533 AA.

AC Q4SYF6;

DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.

DT 19-JUL-2005, sequence version 1.

DT 07-FEB-2006, entry version 4.

DE Chromosome undetermined SCAF12073, whole genome shotgun sequence. (Fragment).

DE ORFNames=GSTENG00010356001;

GN Tetraodon nigroviridis (Green puffer).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

OC NCBI_Taxid=99883;

OX [1]

RN NUCLEOTIDE SEQUENCE.

RX PubMed=15496914; DOI=10.1038/nature03025;

RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N., Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A., Nicaud S., Jaffe D., Fisher S., Luthalla G., Dossat C., Segurens B., Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S., Anthonard V., Jubin C., Castellio V., Katinka M., Vacherie B., Blemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V., Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J., Parra G., Lardier G., Chappell C., McKernan K.J., McEwan P., Bosak S., Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J., Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M., Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C., Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;

RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype."

RT Nature 431:946-957(2004).

RN [2]

RN NUCLEOTIDE SEQUENCE.

RG Genoscope; Whitehead Institute Centre for Genome Research; RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

CC -! CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; CAEE01012073; CAF94326.1; -; Genomic_DNA.

DR NON TER 1 1

FT NON TER 533 533

SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;

Query Match 46.1%; Score 1289; DB 2; Length 533;

Best Local Similarity 48.2%; Pred. No. 1.5e-88;

Matches 257; Conservative

Qy 41 LRSVSPFLITDASLATDPRLFTFLGSPRLRLARGLSPAYLRFGTKTDLFIQDPDK 100

Db 1 LRRVDPFLSVTTDASLAADERFWYLSSPKVRTLAKALPAFLRFGTQDFWVFAPHK 60

Qy 101 -EPTS--EERSYKWSQVNHDCIRSEPVSAVLRLQVWPPFQELLRLREQVQKEFQNSTY 157

Db 61 NQPASGFSARELIFSNGQSCSKMAPPWLERLRRLTEWKKQVWLRLNEELQRVRRVKF 120

Qy 158 SRSSVDMLYSFAKSGGLDLIFGLNALLRTDPDLRWNSNAQLLLDYCSKGVNISELQNE 217

Db 121 TETTVDLQHAFANCSGLDLVFLGNALLRTADNRWNSNARSLLRYCBAARYHMSWELGNE 180

Qy 218 PNSFWKKAHILDLGLQIGEDFVELHKLQRSAP-QNAKLYGPDIGQPGKTKVLLRSLK 276

Db 181 PNSYKEAGRLDGRGLQEDFTVLRLKRSRFRDAGLFGDVGQDPKDRHIDILSGFLQ 240

```
QY 277 AGGEVIDSTWTHYYLNGRIATKEDFLSSDALDTFILSVQILKLVTKETIPGKKVWLGET 336
Db 241 SGAEEVADCTWTHYYLDGREASLEDFLDPOVDTLREKIGEVLEEVHQVSPGKPVWLGET 300
QY 337 SSAYGGGAPLISNTFAAGFMWLDKLGLSAQMGLGVNVRQVFFGAGNHYLVNDFEPLP-- 394
Db 301 SSAYGGGAAGLSDTFFVAGFMWLDKLGLAATLGLVNMVRLQVLIGAGSYHMDNDLPLPRS 360
QY 395 -----DYMLSLFLFKLVGPVRLSR-VKGPDRS-KLRVYLHCTN----- 431
Db 361 GLLQLDYWLSLLYKRLVQGVLEKTRHTPPGASERVLYLHCANKQRCSSLLQFLSVRKQR 420
QY 432 -----VYHPRYQEGDLTVLYNLNHNVTYKHLKVPPLFRKPVDTYLLKPS--GPDGLLS 482
Db 421 KEARFSLVSLCSYRSGAATLSMNLKQPARISLPRILSSSTVEAFVLESEQPGEGELRS 480
QY 483 KSVQLNGQILKMWDEQTLPALTEKPLPAGSALSILPAFSYGFVIRNAKIAACI 535
Db 481 RAVKLNIGRVLRMVDDTFPELEGRSLPAEHLQLPAYSLAFFVFTDAQAAGCV 533

RESULT 12
HPSE2_HUMAN STANDARD; PRT; 592 AA.
AC Q8WWQ2; Q5VUH4; Q5VUH6; Q8WWQ1; Q9HB37; Q9HB38; Q9HB39;
DT 25-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 25-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 16.
DE Heparanase-2 (EC 3.2.-.-) (Hpa2).
GN Name=HPSE2; Synonyms=HPA2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
RX MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbrc.2000.3586;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and expression profiling of Hpa2, a novel mammalian
RT heparanase family member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177 (2000).
[2]
RN NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
[3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TYR-579.
RX PubMed=15164054; DOI=10.1038/nature02462;
RA Deloukas P., Bartholomew M.E., Grafham D.V., Rubinfeld M., French L.,
RA Steward C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,
RA Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.L.,
RA Taylor A., Batties J., Bird C.P., Ainscough R., Almeida J.P.,
RA Ashwell R.I.S., Ambrose K.D., Babbage A.K., Bagguley C.L., Bailey J.,
RA Banerjee R., Bates K., Beasley H., Bray-Allen S., Brown A.J.,
RA Brown J.V., Burford D.C., Burrill W., Burton J., Cahill P., Camire D.,
RA Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Corby N., Coulson A., Dhumi P., Dutta I., Dunn M., Faulkner L.,
RA Frankish A., Frankland J.A., Garner P., Garnett J., Griddle S.,
RA Griffiths C., Grocock R., Gustafson E., Hammond S., Harley J.L.,
RA Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.J.,
RA Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,
RA Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,
RA Lee H.M., Leongamornlert D.A., Laird G., Lloyd C., Lloyd D.M.,
RA Loveland J., Lovell J., McLaren S., McLay K.E., McMurray A.,
RA Mashreghi-Mohammadi M., Matthews L., Milne S., Nickerson T.,
RA Nguyen M., Overton-Larty E., Palmer S.A., Pearce A.V., Peck A.I.,
RA Pelan S., Phillimore B., Porter K., Rice C.M., Rogosin A., Ross M.T.,
```

```
RA Sarafidou T., Sehra H.K., Showkeen R., Skuce C.D., Smith M.,
RA Stranding L., Sycamore N., Tester J., Thorpe A., Torcasso W.,
RA Tracey A., Tromans A., Tsolas J., Wall M., Walsh J., Wang H.,
RA Weinstock K., West A.P., Willey D.L., Whitehead S.L., Wilming L.,
RA Wray P.W., Young L., Chen Y., Lovering R.C., Moschonas N.K.,
RA Siebert R., Fichtel K., Bentley D., Durbin R., Hubbard T.,
RA Doucette-Stamm L., Beck S., Smith D.R., Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 10.";
RL Nature 429:375-381 (2004).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -!- SUBCELLULAR LOCATION: Membrane-associated.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=HPA2c;
CC IsoId=Q8WWQ2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8WWQ2-2; Sequence=VSP_015852, VSP_015853;
CC Name=3; Synonyms=HPA2b;
CC IsoId=Q8WWQ2-3; Sequence=VSP_015851;
CC Name=4; Synonyms=HPA2a;
CC IsoId=Q8WWQ2-4; Sequence=VSP_015850;
CC -!- TISSUE SPECIFICITY: Widely expressed, with the highest expression
CC in brain, mammary gland, prostate, small intestine, testis and
CC uterus. Found both in normal and cancer tissues.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AF282885; AAG23421.1; -; mRNA.
DR EMBL; AF282886; AAG23422.1; -; mRNA.
DR EMBL; AF282887; AAG23423.1; -; mRNA.
DR EMBL; AJ299719; CAC82491.1; -; mRNA.
DR EMBL; AJ299720; CAC82492.1; -; mRNA.
DR EMBL; AL590036; CAH73137.1; -; Genomic DNA.
DR EMBL; AL139243; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL445251; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL139243; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH73137.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH70448.1; -; Genomic DNA.
DR EMBL; AL139243; CAH70448.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH70448.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH70448.1; JOINED; Genomic DNA.
DR EMBL; AL445251; CAH70448.1; JOINED; Genomic DNA.
DR EMBL; AL590036; CAH70448.1; JOINED; Genomic DNA.
DR EMBL; AL445251; CAH16472.1; -; Genomic DNA.
DR EMBL; AL139243; CAH16472.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH16472.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH16472.1; JOINED; Genomic DNA.
DR EMBL; AL590036; CAH16472.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH17160.1; -; Genomic DNA.
DR EMBL; AL356268; CAH17160.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH17160.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH17160.1; JOINED; Genomic DNA.
DR EMBL; AL590036; CAH17160.1; JOINED; Genomic DNA.
DR EMBL; AL139243; CAH73139.1; -; Genomic DNA.
DR EMBL; AL356220; CAH73139.1; JOINED; Genomic DNA.
DR EMBL; AL356268; CAH73139.1; JOINED; Genomic DNA.
DR EMBL; AL445251; CAH73139.1; JOINED; Genomic DNA.
DR EMBL; AL356220; CAH17162.1; -; Genomic DNA.
DR EMBL; AL356268; CAH17162.1; JOINED; Genomic DNA.
DR EMBL; AL139243; CAH17162.1; JOINED; Genomic DNA.
```


GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.8121 Seconds
(without alignments)
2889.939 Million cell updates/sec

Title: US-10-645-659A-2
Perfect score: 2797
Sequence: 1 MLRLLLLMWPLGALQAQA.....LPAPSYGFFVIRNAKIAACI 535

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.5	32.0	480	JC7506	heparanase protein
2	391	14.0	521	T45608	hypothetical prote
3	161.5	5.8	190	T01953	hypothetical prote
4	117	4.2	1331	A48954	mannan endo-1,4-be
5	114.5	4.1	1573	T50113	3-dehydroquinatase
6	113.5	4.1	837	A31842	endo-1,4-beta-xyla
7	111.5	4.0	599	S75363	hypothetical prote
8	110	3.9	575	T45668	hypothetical prote
9	109.5	3.9	2165	RRN2A2	genome polypeptid
10	105.5	3.8	897	G02529	dynein heavy chain
11	105.5	3.8	464	A38905	dynein heavy chain
12	104.5	3.7	587	S36231	beta-fructofuranos
13	104	3.7	484	G72395	alpha-L-arabinofur
14	103.5	3.7	960	T37916	probable heterochr
15	103	3.7	464	C40630	GDP-mannose pyroph
16	103	3.7	500	D87541	beta-xylosidase [i
17	103	3.7	557	ODNC1	cytochrome-c oxida
18	102.5	3.7	505	G86391	hypothetical prote
19	102	3.6	596	T04506	ATP-dependent nucl
20	102	3.6	1180	A13504	cytochrome-c oxida
21	101.5	3.6	523	T11317	formin-binding pro
22	101.5	3.6	844	T52396	DNA topoisomerase
23	101.5	3.6	1462	T06919	hypothetical prote
24	100.5	3.6	335	T05268	hypothetical prote
25	100.5	3.6	356	D85406	hypothetical prote
26	100	3.6	714	CIU0H	calpain (EC 3.4.22
27	99.5	3.6	1392	YGBYAD	L-aminoadipate-sem
28	99	3.5	709	A35364	carcinoembryonic a
29	99	3.5	824	T10615	hypothetical prote

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human
C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircok, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

C:Comment: This protein, an intracellular membrane-bound enzyme, has biological and therape

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match	32.0%	Score	895.5	DB 2	Length	480			
Best Local Similarity	37.9%	Pred. No.	1.1e-61						
Matches	194	Conservative	70	Mismatches	141	Indels	107	Gaps	6
Qy	33	LEPYTKRLRSPVSPFLSITIDASLATDPRLFTFLGSPRLRALARGLSPAYLRFGGKTKTD	92						
Db	63	LDVSTKNPVRTVNVNENFLSLQDPSIIHD-CWLDLFLSSKRLVTLARGLSAPFLRFGGKRTD	121						
Qy	93	FLIFDPPDKPTSEERSYWKSVQNHIDICRSPVSAAVLRKLQVWPPFOELLLRLRQYQKEF	152						
Db	122	FLQFQNLNRPAS-----RGPGPF-----DYLKNY	147						
Qy	153	KNSTYSSRSVDMLYSFAKCSGLDLIFGLNALLRTPDLRWNSSNAQLLLDYCSSKGYNISW	212						
Db	148	E-----	148						
Qy	213	ELGNPNPSFWKKAHLIDGLQLGDFEVLHKLQOR-SAFQNAKLYGPDIGQPRGKTVKLL	271						
Db	149	---DEPNRYTMHGRAVNSQLGDKYIQLKSLQPIRYSRASLYGNIGRPKNVIAL	205						
Qy	272	RSFLKAGGEVIDSLTWHHYLYNGRIATKDEFLSSDALDTPILSVQKILKVTKEITPGKKV	331						
Db	206	DGFMKVGAGTVDATWQHICYIDGRVVKVMDFLKTRLLDTSQIRKIQKVNTYTPGKKI	265						
Qy	332	WLGETSSAYGGGAPLLSNTPAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDENFE	391						
Db	266	WLEGVVTTSAGGTNNLSYAAAGFLWNLTLGLMANQGIQDVVIRHSFDPDGHYNHLVDQNFN	325						
Qy	392	PLPDYWLSLLFKKLVGPRVLLSRVKGPD-----RSKLRVYLHCTNVYHPRVOEGDL	442						
Db	326	PLPDYWLSLLYKRLTIGPKVLAVHVGAGLQRPGRVIRDKRIYAHCTNNHHNNYVRGSI	385						

	Matches	108;	Conservative	72;	Mismatches	194;	Indels	192;	Gaps	24;	
Qy	21	PAGTA-PTDDVVLDLBEFYTKRPLRSVSPSFLSITIDASLATDPRFLTFLGSLRLRAL---	A	76							
Db	759	FTVTATPTPTPIPTVTPLPITISPSVVEITINTNAGRTOI-----SPIYGANODI	812								
Qy	77	RGLSPAYLRFGGTK-----TDFLI PDPDEKPTSEERSYKSKOVNHDI CRSEPVAAVL	129								
Db	813	EGVHSAARLGNGRLTYGNWENNFSNACNDWYHSSDDYLCSMGISGEDAK---VPAAVV	869								
Qy	130	RKLQVWFQBELLLLRLEQYKEFKNSTYSRSSVDML-----	165								
Db	870	SKF-----HEVSLKNNAVSATLQWAGTVSKDNYGVSTENETAFNSRWAE	914								
Qy	166	YSFAKCGSLDLIFGLNALLRTPDLRNWSNAQLLLDYCSSKGYNIS-----WELGNE	217								
Db	915	VKFKKDAPLSL-----NPDLNDFVMDEFINYLINK-YGMASSPTGIKGIVILDNE	964								
Qy	218	PNSFWKKAHIILDG-----LOIGEDFVELHLKLLORSAPONAKLYG-----	257								
Db	965	PD-LWASTHPRIHPNKVTCNELIEKSVELAKVI-KTLDPESAEEVFASYGFMGYSLQDA	1022								
Qy	258	PDIQOPRG-----KTVKLLRSFLKAGGEVIDSLTWHHVYLVNGRIATKESDFLSSD	306								
Db	1023	PDMNQVGHRWFISWYLEQMKKASDSFGKRLDVDL-LHWYPARGGNIRVCFDGENDT	1081								
Qy	307	ALDTFILSVQK-----ILKVTKKITPGKKVLMGETSYAAGGGAPLLSNT-----	350								
Db	1082	SKEVVIARMQAPRTLWDPTYKTSVKGQITAGENSWINQWFSDY---LPIIPNVKADIEKY	1138								
Qy	351	-----FAAGFWMLDKLGLSAOMGIEVNM-----QVF----	377								
Db	1139	YPGTKLAISEPDYGGERNHISGIALADVIGFYGVNFARWGDSGSAAAAANYILNY	1198								
Qy	378	-----FGAGNYHLVDENFEPLDPLWLSLLPKVLGVRVRLSRVKGPDRSKLRVLHCTN	431								
Db	1199	DGKSGKYGNTVNSANTSDVENMPVY-----ASINGQDDSELHILLNRN	1242								
Qy	432	VYHPRYQEGDLTYLVNLHNVTXHLK	457								
Db	1243	-YDQKLQ-----VKINITSPKYTK	1261								
	RESULT	5									
	T50113	3-dehydroquininate synthase (EC 4.2.3.4) - fission yeast (Schizosaccharomyces pombe)									
N:	Contains:	3-dehydroquininate dehydratase; 3-dehydroquininate synthase (EC 4.6.1.3)									
C:	Species:	Schizosaccharomyces pombe									
C:	Date:	09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004									
R:	Accession:	T50113									
R:	Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.										
A:	submitted to the EMBL Data Library, February 2000										
A:	Reference number:	Z25039									
A:	Accession:	T50113									
A:	Status:	preliminary; translated from GB/EMBL/DBJ									
A:	Molecule type:	DNA									
A:	Residues:	1-1573 <SEE>									
A:	Cross-references:	UNIPROT:Q9P7R0; UNIPARC:UPI0000125F22; EMBL:AL157734; PIDN:CAB75770									
A:	Experimental source:	strain 972h(-); cosmid c1834									
C:	Genetics:										
A:	Gene:	SPDB:SPAC1834.02									
A:	Map position:	1									
C:	Superfamily:	pentafunctional Arom protein; 3-dehydroquininate dehydratase homology; 3-dehydroquininate kinase homology									
C:	Keywords:	carbon-oxygen lyase; phosphorus-oxygen lyase									
F:	407-835/Domain:	3-phosphoshikimate 1-carboxyvinyitransferase homology <PSK>									
F:	1035-1279/Domain:	3-dehydroquininate dehydratase homology <DQD>									
	Query Match	4.1%; Score 114.5; DB 2; Length 1573;									
	Best Local Similarity	21.1%; Pred No. 3.8;									
	Matches	120; Conservative	83; Mismatches	184; Indels	183; Gaps	32;					
Qy	18	QGAPAGTAPTDVVVDLEFYTKRPLRSVSPSFLSITIDASLAT-----DP-RFLTFLGSPR	71								

Db	699	QGPPKGTGLKPLESIDME-----TWDAFLTASVVAACVNSVSGDPVTRITGIANQR	750
Qy	72	LR-----ALARGLSPAYLRFGGTKDFLFLDPD-----KEPTSEERSYKSNQVNHDCRS	121
Db	751	VKECNRIAAVMVHELAKFGVKTGLEDTGYIFGKNYKELKKP---EEGIY---TYDDHRIAMS	806
Qy	122	E-----PVSAAVLRKLQVE---WPFQELLLREOYQKEPKFNSTYSSRSDMLYSFAKCS	172
Db	807	FSVLSLICPSTLLIIDKACVEKTPYW-WDVLHQSGVGLTGAT-SVASDPLKGSISKNA	864
Qy	173	GLDLIFGLNALLTP-----DLRWSSNAQLLLDYCSKSGYNISWELGNPEP	218
Db	865	SIILI-GMRGAGKTTIGKIIFAKQLNPKFLDL-----DELLEDYLEMPIAEVIFRMG---	914
Qy	219	NSFWKKAHILGLQGEDFVELHKLORSAFONAKLYGPDIGOPGRKTVK-----LLR	272
Db	915	--W-----DAFRLEE-----HKVLKPIETHPEGY---VAASGGVVIEMDESNNLLS	956
Qy	273	SFLKAGEVIDSLTWHYYLNGRIATKEDFLSSDALDFTILSVQILKVTKETPGKKVW	332
Db	957	NFKVKEGGIVL-----HVHRN---LEHIKSYLSQDQTRPTKYDQESIDDDVYKR---	1004
Qy	333	LGSTSAAYGGGAPILSNTPAAGFMWLDKLGLSAQMGIEMVMQVFGAGNYHLVDENFEP	392
Db	1005	YRECRSHY-FISPVLSN-----QVIDEKIQ-1028	
Qy	393	LPDYWLSLLFKVLGPRVLLSRVKGDPDRSKLRVYLHCTNVMHPRYOEGDLTVLVNLHNV	452
Db	1029	---YSRSRFLDVVTGSSQVLQFKTKKRSTF-----LTLNVPRIEDALPTL-----	1071
Qy	453	TKHLKVPVPLFRKPVDVTVLLKPSGPDGLLSKSVQLNGQILKMVDEQ-----TLPAL-	503
Db	1072	RDVTVGCDIAEVRVD-YLKDPKSSNGISS-----LDFVAEQISLLRCSTTLLPIIF	1120
Qy	504	TEKPLPAG-----SALSPLAPSVG	522
Db	1121	TIRTISQGLFPNDKEEAEKELMSAMRYG	1150
RESULT 6			
A31842			
Endo-1,4-beta-xylanase (EC 3.2.1.8) Z precursor - Clostridium thermocellum			
N/Alternate names: xylanase Z			
C/Species: Clostridium thermocellum			
C/Date: 31-Mar-1990 #sequence_revision 11-Apr-1997 #text_change 09-Jul-2004			
C/Accession: A31842			
R/Greipinet, O.; Chebrou, M.C.; Beguin, P.			
J. Bacteriol. 170, 4582-4588, 1988			
A/Title: Nucleotide sequence and deletion analysis of the xylanase gene (xynZ)			
A/Reference number: A31842; MUID:89008072; PMID:3139632			
A/Accession: A31842			
A/Molecule type: DNA			
A/Residues: 1-837 <GRE>			
A/Cross-references: UNIPROT:P10478; UNIPARC:UPI000013909C; GB:M22624; NID:G14493			
C/Genetics:			
A/Gene: xynZ			
C/Function:			
A/Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans			
A/Pathway: xylan degradation			
C/Superfamily: Clostridium endo-1,4-beta-xylanase Z; Clostridium cellulase repeat			
C/Keywords: duplication; extracellular protein; glycosidase; heat-stable protein			
F/1-28/Domain: signal sequence #status predicted <SIG>			
F/29-837/Product: endo-1,4-beta-xylanase #status predicted <MAT>			
F/326-419/Domain: Clostridium xylanase A repeat homology <XA>			
F/430-453/Domain: Clostridium cellulase repeat homology <CCR1>			
F/464-487/Domain: Clostridium cellulase repeat homology <CCR2>			
F/548-834/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>			
F/645,754/Active site: Glu #status predicted			
Query Match 4.1%; Score 113.5; DB 1; Length 837;			
Best Local Similarity 19.5%; Pred. No. 1.8;			
Matches 85; Conservative 66; Mismatches 155; Indels 129; Gaps 20;			

Db 271 -----TKEG--PPDAKMFVLSVSKILYVGM-----KKGNFQHSFLAGGA 310
Qy 345 PLLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVPGAGNYHLVDENRPPDPYWLSSLFKK 404
Db 311 -----TLSAGRIYVD-----DGLKAVWPHSGHYLPTEENFOA-----FMSFLREN 351
Qy 405 LVGPRVLLSRVKGPD 419
Db 352 NVD---LANVKNPD 363

RESULT 9
RRNZ2

genome polyprotein - human respiratory syncytial virus (strain A2)
N:Alternate names: polymerase L protein
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: human respiratory syncytial virus
A:Note: host Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 09-Jul-2004
C:Accession: A40317; A28319; PS0048
R:Stec, D.S.; Hill III, M.G.; Collins, P.L.
Virolgy 183, 273-287, 1991
A:Title: Sequence analysis of the polymerase L gene of human respiratory syncytial virus
A:Reference number: A40317; MUID:91272488; PMID:2053282
A:Accession: A40317
A:Molecule type: mRNA
A:Residues: 1-2165 <STE>
A:Cross-references: UNIPROT:P28887; UNIPARC:UPI0000134AEA; GB:M75730; NID:G333955; PIDN:
R:Collins, P.L.; Olmsted, R.A.; Spriggs, M.K.; Johnson, P.R.; Buckler-White, A.J.
Proc. Natl. Acad. Sci. U.S.A. 84, 5134-5138, 1987
A:Title: Gene overlap and site-specific attenuation of transcription of the viral polyome
A:Reference number: A28319; MUID:97260943; PMID:2440043
A:Accession: A28319
A:Molecule type: DNA
A:Status: preliminary; not compared with conceptual translation
A:Residues: 1-81 <COL>
A:Cross-references: UNIPARC:UPI0000134AE9; GB:M17245; NID:G333953; PIDN:AAA47417.1; PID:
R:Johnson, P.R.; Collins, P.L.
J. Gen. Virol. 69, 2901-2906, 1988
A:Title: The A and B subgroups of human respiratory syncytial virus: comparison of inter
A:Reference number: PS0048; MUID:89036169; PMID:3183631
A:Accession: PS0048
A:Molecule type: mRNA
A:Residues: 1-18 <JOH>
A:Cross-references: UNIPARC:UPI0000172722; GB:D00397; NID:G222551; PID:G2160375
A:Experimental source: strain 18537
A:Note: this strain belongs to subgroup B
C:Genetics:
A:Gene: L
C:Superfamily: parainfluenza virus RNA-directed RNA polymerase
C:Keywords: ATP; nucleotidyltransferase

Query Match 3.9%; Score 109.5; DB 1; Length 2165;
Best Local Similarity 22.5%; Pred. No. 15;
Matches 89; Conservative 48; Mismatches 119; Indels 139; Gaps 21;
Qy 226 HILIDGLQGEDFVELHKLQSAFONAKLYGPD-IGQPRGKTVKLSRSLKAGGEVIDS 284
Db 1420 LLMKPPIFTGD--VDIHKLQ--VIOQHMFDPKISLTQYVELFUSNKTLSGSHVNSN 1475
Qy 285 LTHWH---YYLNGRIATKEDFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAY 340
Db 1476 LILAHKISDYFHTYI-----LSTNLAGHWILLIQ-LMKDSKGF--EKDW-GE----- 1520
Qy 341 GGGAPLLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLV-----D 387
Db 1521 -----GY-----ITDHMFNL---KVFFNAYKTYLLCFHKGYKAKLECD 1557
Qy 388 ENFEPL-----PDWLSL-----LFPKLVGRVLLSRVKGPDRLKRLVYLHCTN 431
Db 1558 MNTSDLLCVLELIDSSYWKMSKSVFLEQKVIKYSODASLHRVKGCHSPKL-WFLKRLN 1616
Qy 432 V-----YHPRYQEGDLT-----LYVLNL--- 449

Db 1617 VAEFTVCPVVVNIYHPTMKAILTYIDLVRMGLINIDRIHIKNKHKNFDEFTSNLFYI 1676
Qy 450 -----HNVTKHLKP-PPLFRKPVDTYLLKPSGPDGLLSKSVQLNGQ-----ILK 493
Db 1677 NYNFSDNTHLTUKHIRANSLENNYNKLYHPTPETLENILANPIKSNDDKTLNDYCIQK 1736
Qy 494 MVDEQTLPALTEKPLPAGSALSPLAFS-----YGFF 524
Db 1737 NVDSIMLPLLSNKKLIKSSAMIRTYNSKQDLYNLF 1771

RESULT 10

G02529
dynein heavy chain 1, cytosolic - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
R:Accession: G02529
R:Vaisberg, E.A.; Griesom, P.M.; McIntosh, J.R.
submitted to the EMBL Data Library, April 1996
A:Reference number: H01399
A:Accession: G02529
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-897 <VAL>
A:Cross-references: UNIPROT:Q14204; UNIPARC:UPI000016B1B7; EMBL:U53530; NID:G1314642; PI
C:Superfamily: dynein heavy chain, cytosolic
Query Match 3.8%; Score 105.5; DB 2; Length 897;
Best Local Similarity 19.3%; Pred. No. 8.2;
Matches 116; Conservative 85; Mismatches 206; Indels 195; Gaps 26;
Qy 38 KRPLRSVSPSLSTIDASLATDPRFLTFLGSPRLRALARGLSPAYLRGGTKTDLIFD 97
Db 219 EQPWVSQPRKLRQNLDAALLQLKSF-----PARLRQYASVEFVORLLKGYMKINMLVIE 273
Qy 98 PDKEPTSEERSYWKSVQNVNDICRSEPVSAALVKLQVWPPFOELL----- 143
Db 274 LKSEALKDR--HWKO-----LMKRLHVNWVWVSELTGQIWDVDLQKNEAI 316
Qy 144 -----LREQYQKEPKN--STYSRSVDMLYSFAKCSGLDLPGLNALLRTPDL 189
Db 317 VKDVLLVAQGEAMALEEFKQIREVWNTY---ELDLVNYQNKCR---LIRGWDDLFNKVK 370
Qy 190 RWNSSNAQLLDYCSSKGYNI-----SWEIGNEPN-----SFW---KKAHILIDGLQ 234
Db 371 HINSVSAMKLSPY-----YKVFEDALSWE--DKLNRIMALFDWIDVQRRWVYLEGIFT 423
Qy 235 GE-----DFVELHKLQSAFONAKLYGPDIGQPRGKTVKLSRSL 275
Db 424 GSADIKHLLPVETORFQSI STEFLAMKKVSKSLVMDVLNTQGVQSRLESLADLLGKITQ 483
Qy 276 KAGEVI--DSLTHWHYVLANGRIATKEDFLSSDALDFTILSVQKILKVTKEITPGKKVWL 333
Db 484 KALGEYLERERSSPRFYFVG-----DEDLLE-----IIGNSKNVAKLQKH--KMF 530
Qy 334 GETSSAYGGGAPLLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPL 393
Db 531 GVSSIIILNEDNSV-----LGISREGEVWFKTP-----VSITEHPK 568
Qy 394 PDYWLSSLFPKLVGRVLLSRVKGPDRLKRLVYLHCTN-----YHPRYQ----- 439
Db 569 INEWLTLVEKEM--RVTLAKLAESVTEVEIFGKATSIDPNTYITWIDKYQALVLSA 625
Qy 440 -----GDLTLVNLHNTYKHLKY-----PPPLFRKPEVD----- 468
Db 626 QIAWSENVTALSSMGGGDAAPSDSVLSNVEVTLNVLADSVLMEQPPPLRRRLKHLITE 685
Qy 469 -----TYLLKPSGPDGLLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFS 521
Db 686 LVHORDVTRSLIKSIDN--AKSFEWLSQMRPFYDPKQTDVLQQLSIQWANA-----KFNY 739
Qy 522 GF 523

Db 740 GF 741

RESULT 11

A38905

dynein heavy chain, cytosolic - rat

N;Contains: dynein ATPase (EC 3.6.4.2)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 15-Apr-1994 #sequence_revision 02-May-1994 #text_change 09-Jul-2004

C;Accession: A38905; I58139

R;Zhang, Z.; Tanaka, Y.; Nonaka, S.; Aizawa, H.; Kawasaki, H.; Nakata, T.; Hirokawa, N. Proc. Natl. Acad. Sci. U.S.A. 90, 7928-7932, 1993

A;Title: The primary structure of rat brain (cytoplasmic) dynein heavy chain, a cytoplasmic dynein heavy chain

A;Reference number: A38905; MUID:93376715; PMID:7690137

A;Accession: A38905

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-4644 <ZHA>

A;Cross-references: UNIPROT:P38650; UNIPARC:UPI000013C4AA; GB:D13896; NID:g402527; PIDN:R1Mikami, A.; Paschal, B.M.; Mazumdar, M.; Vallee, R.B. Neuron 10, 787-796, 1993

A;Title: Molecular cloning of the retrograde transport motor cytoplasmic dynein MAP 1C.

A;Reference number: I58139; MUID:93264075; PMID:7684232

A;Accession: I58139

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1023, 'MP', 1026-1771, 'D', 1773-2097, 'A', 2099-2138, 'V', 2140-2174, 'A', 2176-2184 <RES>

A;Cross-references: UNIPARC:UPI000016867D; GB:I08505; NID:g294542; PIDN:AAA41103.1; PID:C;Superfamily: dynein heavy chain, cytosolic

C;Keywords: ATP; blocked amino end; heterotrimer; hydrolase; microtubule binding; nucleotide-binding motif A (P-loop)

F:1904-1911/Region: nucleotide-binding motif A (P-loop)

F:2222-2229/Region: nucleotide-binding motif A (P-loop)

F:2593-2600/Region: nucleotide-binding motif A (P-loop)

F:2935-2942/Region: nucleotide-binding motif A (P-loop)

F:1910/Binding site: ATP (lys) #status predicted

F:2228/Binding site: ATP (lys) #status predicted

F:2599/Binding site: ATP (lys) #status predicted

F:2941/Binding site: ATP (lys) #status predicted

Query Match 3.8%; Score 105.5; DB 1; Length 4644;

Best Local Similarity 19.3%; Pred. No. 96;

Matches 116; Conservative 84; Mismatches 207; Indels 195; Gaps 26;

Qy 38 KPRLRSVSFLSITIDASLATDPRFLTFLGSLRALARGLSPAYLRFQGTDTDFLIPD 97

Db 1346 EQPWSVQPKRLQNLQDLNQLNKF-----PARLRQYASYEFVQRLKKGMYKINMLVIE 1400

Qy 98 PDKEPTSEERSYKMSQVNHDCRSEPVSAAVLRKLOVEWPFQELLL----- 143

Db 1401 LKSEALKDR--HWKQ-----LMKRLHNVWVSELTGQIWDVDLQKNEAI 1443

Qy 144 -----LREOYQKEFN--STYSSSSVDMLYSPAKCSGLDLIFGLNALLRTPDL 189

Db 1444 VKDVLVAQGEALEBFLQIREVNTY---ELDLVNYQNKCR---LIRGWDLLFNKVK 1497

Qy 190 RWNSSNAQLLLDYCSSKGYNI-----SWELGNEPN-----SFW---KKAHLIDGLQL 234

Db 1498 HINSVSAMKLSPY-----YKVPFEDALSWE--DKLARIMALFDVWIDVQRRWYLEGIFT 1550

Qy 235 GE-----DVELHKLQRFSAFNQAKLYGPDIGOPRGKTVKLLRSPL 275

Db 1551 GSADIKHLPLVETQRFQSTSTEFALMKRVSKPLVMDVLNIGOVRSRLRDLGLKIQ 1610

Qy 276 KAGGEVI--DSLFWHHYILNGRIATKDEFLSSDALDTFILSVOKILKVTKEITPGKKVWL 333

Db 1611 KALGEVLERSERSFPFYVG-----DEDLLE-----IIGNSKNVAKLQKHF---KKMFA 1657

Qy 334 GETSSAYGGAPLLSNTFFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNTHLVDENPFL 393

Db 1658 GVSSIIILNDSVV-----LGTISREGSEVMPKTP-----VSITEHPK 1695

Qy 394 PDYWLSSLFFKLVGPRVLLSRVKGPDRSKRLRVYLHCT-----NVYHPRVOEGDLTYVLN- 448

Db 1696 INEWLTLVEKEM---RVTAKLLAESVTEVEIFGKATSIDENTYITWIDKYOQLWLSA 1752

Qy 449 -----LHNVTYKHLKV-----PPPLFRKPVD----- 468

Db 1753 QIAWSENVENALSNVGGGNGVGPLOSVLNSVEVTLNVLADSVLMQPPFLRRKLEHLITE 1812

Qy 469 -----TYLLKPSGPDGLLSKSVQLNGQILKMWDEOTLPALTEKPLPAGSALSLPATSY 521

Db 1813 LVHQRDVTRSLIKSIDN--AKSFELWSQRFYDFDPKQTDVLQQLSIQMANA-----KFN 1866

Qy 522 GF 523

Db 1867 GF 1868

RESULT 12

S36231

beta-fructofuranosidase (EC 3.2.1.26) - potato (fragment)

N;Alternate names: invertase

C;Species: Solanum tuberosum (potato)

C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004

C;Accession: S36231; S34145

R;Hedley, P.E.; Machray, G.C.; Davies, H.V.; Burch, L.; Waugh, R. Plant Mol. Biol. 22, 917-922, 1993

A;Title: cDNA cloning and expression of a potato (Solanum tuberosum) invertase.

A;Reference number: S36231; MUID:93363925; PMID:8358038

A;Accession: S36231

A;Molecule type: mRNA

A;Residues: 1-587 <HED>

A;Cross-references: UNIPROT:Q43172; UNIPARC:UPI0000175B36; EMBL:Z21486

R;Hedley, P.E.; Machray, G.C.; Davies, H.V.; Burch, L.; Waugh, R. submitted to the EMBL Data Library, January 1993

A;Reference number: S34145

A;Accession: S34145

A;Molecule type: mRNA

A;Residues: 1-576, 'KM', 579-587 <HED>

A;Cross-references: UNIPARC:UPI0000AD071; EMBL:Z21486; NID:g313128; PIDN:CAA79676.1; PIIC;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

Query Match 3.7%; Score 104.5; DB 2; Length 587;

Best Local Similarity 26.2%; Pred. No. 5.2;

Matches 43; Conservative 22; Mismatches 54; Indels 45; Gaps 8;

Qy 77 RGLSPAYLREGGKTDTFLIFDPDKEPTSEERSYK-----KSQVNHDCRSEPVSAAVLRK 131

Db 306 KGLRLDYNTYASKS---FYDPSK---NRRIMMGWANESDTVNDVKKGWAGIQTIPRK 358

Qy 132 L-----QVEWPFQELLLREOYQK-----EFKNSTYSSSSVDMLYSPAKC 171

Db 359 LWDPSKGQVLQVQVVELETLREQKQVLSNRKLLKGGKIEVKGITPAQADVETFSF--- 415

Qy 172 SGLDLIFGLNALLRTPDLRWNSSNAQLLLDYCSSKGYNISWELG 215

Db 416 SSLD-----KAEPFDPNDNLVQ---DVCAIKGSTVQGDIG 449

RESULT 13

G72395

alpha-L-arabinofuranosidase - Thermotoga maritima (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

C;Accession: G72395

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: G72395

A;Status: preliminary

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model
Run on: June 5, 2006, 12:01:06 ; Search time 106.661 Seconds
(without alignments)
2293.354 Million cell updates/sec

Title: US-10-645-659A-2
Perfect score: 2797
Sequence: 1 MLRLLLWLWGLGALAQA.....LPAPSYGFFVIRNAKIAACI 535

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*
10: Geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2797	100.0	535	3 AAB08851	Aab08851 A murine
2	2797	100.0	535	5 ABB07811	Abb07811 Mouse hep
3	2797	100.0	535	7 ADG88834	Adg88834 Mouse hpa
4	2797	100.0	535	8 ADL16413	Adl16413 Mouse hep
5	2797	100.0	535	8 ADM48750	Adm48750 Mouse hpa
6	2797	100.0	535	8 ADR88208	Adr88208 Mouse hep
7	2797	100.0	535	8 ADT78175	Adt78175 Mouse hep
8	2797	100.0	535	9 AEA42424	Aea42424 Mouse hep
9	2793	99.9	535	9 ADY27033	Ady27033 Murine he
10	2590.5	92.6	536	5 ABB07812	Abb07812 Rat hepar
11	2590.5	92.6	536	8 ADR88209	Adr88209 Rat hepar
12	2590.5	92.6	536	8 ADT78176	Adt78176 Rat hepar
13	2590.5	92.6	536	9 ADY27035	Ady27035 Rat hepar
14	2590.5	92.6	536	9 AEA42425	Aea42425 Rat hepar
15	2184	78.1	545	3 ADY27034	Ady27034 Bovine he
16	2149	76.8	543	2 AAY17082	Aay17082 Human hep
17	2149	76.8	543	4 AAB86206	Aab86206 Human hep
18	2149	76.8	543	7 ADD18950	Add18950 Human dis
19	2149	76.8	543	8 ADK52086	Adk52086 Human ato
20	2149	76.8	543	8 ADM48759	Adm48759 Human hpa
21	2149	76.8	543	8 ADN05074	Adn05074 Antipsori
22	2149	76.8	543	8 ADN04902	Adn04902 Antipsori
23	2149	76.8	543	8 ADQ80372	Adq80372 Heparanas

24	2149	76.8	543	8	ADR88210	Adr88210 Human pre
25	2149	76.8	543	8	ADP25079	Adp25079 PRO poly p
26	2149	76.8	543	8	ADT78177	Adt78177 Human hep
27	2149	76.8	543	9	ADY27036	Ady27036 Human hep
28	2149	76.8	543	9	AEA42426	Aea42426 Human hep
29	2149	76.8	588	2	AAZ30124	Aaz30124 A human p
30	2147.5	76.8	545	6	ABP56822	Abp56822 Human hep
31	2147.5	76.8	545	7	ADE16012	Ade16012 G-coupled
32	2147.5	76.8	545	8	ADL93951	Adl93951 Human G-c
33	2146	76.7	543	2	AAZ02345	Aaz02345 A human h
34	2146	76.7	543	3	AAZ57590	Aaz57590 Human hep
35	2146	76.7	543	3	AAB08849	Aab08849 Amino aci
36	2146	76.7	543	3	AAZ52990	Aaz52990 Human hep
37	2146	76.7	543	4	AAZ97635	Aaz97635 Human hep
38	2146	76.7	543	5	ABB07813	Abb07813 Human hep
39	2146	76.7	543	7	ADG88800	Adg88800 Human hpa
40	2146	76.7	543	8	ADL16379	Adl16379 Human hep
41	2146	76.7	543	8	ADM48716	Adm48716 Human hpa
42	2146	76.7	543	9	AEA42466	Aea42466 Human hep
43	2146	76.7	543	10	AE956848	Ae956848 Human hep
44	2146	76.7	592	2	AAZ02346	Aaz02346 A human h
45	2146	76.7	592	3	AAB08850	Aab08850 Amino aci

ALIGNMENTS

RESULT 1
AAB08851
ID AAB08851 standard; protein; 535 AA.
AC AAB08851;
XX
DT 15-JAN-2001 (first entry)
XX
DE A murine heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX
OS Mus sp.
XX
PN WO200052178-A1.
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US003542.
XX
PR 01-MAR-1999; 99US-00258892.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD..
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
DR WPI; 2000-579289/54.
DR N-ESDB; AAA75081.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
XX
CC Claim 22; Page 144-145; 152pp; English.
XX
CC The present sequence represents murine protein with heparanase catalytic
CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
CC particularly in treating tumour, inflammation or autoimmunity.
CC Particularly, the polynucleotide is useful in modulating the
CC bioavailability of heparin-binding growth factors, cellular responses to

CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
CC susceptibility to certain viral and some bacterial and protozoa
CC infections, or disintegration of neurodegenerative plaques. The
CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
CC radiation burns), and in the treatment of angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
CC Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,
CC bacterial or protozoa infections
XX
XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 3; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.9e-262;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLRLLLWLWGPGLGALAQAAGAPAGTAPDDVVDFEYTKRPLRSVSPFLSITIDASLATD 60
DB 1 MLRLLLWLWGPGLGALAQAAGAPAGTAPDDVVDFEYTKRPLRSVSPFLSITIDASLATD 60
QY 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDI 120
DB 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDI 120
QY 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGL 180
DB 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGL 180
QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240
DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240
QY 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRFLKAGGEVIDSLTWHYYLNGRIATKE 300
DB 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRFLKAGGEVIDSLTWHYYLNGRIATKE 300
QY 301 DFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
DB 301 DFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
QY 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFLFKLVGPRVLLSRVKGPD 420
DB 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFLFKLVGPRVLLSRVKGPD 420
QY 421 SKLRVYLHCTNVVHPRYQEGDLTYLVNLHNVTKHLKVPVDPYTLKXSPGPDGL 480
DB 421 SKLRVYLHCTNVVHPRYQEGDLTYLVNLHNVTKHLKVPVDPYTLKXSPGPDGL 480
QY 481 LSKSVQLNGQILKMWVDEQTLTPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535
DB 481 LSKSVQLNGQILKMWVDEQTLTPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535

RESULT 2

ID ABB07811 standard; protein; 535 AA.

XX ABB07811 standard; protein; 535 AA.

AC ABB07811;

DT 03-JUL-2002 (first entry)

XX Mouse heparanase sequence.

DE Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;

XX anti-protozoan; neuroprotective; heparin; mouse.

OS Mus musculus.

XX Key Location/Qualifiers

FT Peptide /note= "putative signal peptide"

FT Protein /note= "mature protein"

Db 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535
|||||
RESULT 3
ADG88834
ID ADG88834 standard; protein; 535 AA.
XX
AC ADG88834;
XX
DT 11-MAR-2004 (first entry)
XX
DE Mouse hpa protein.
XX
KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;
necrosis; pressure wound; diabetic ulcer; angiogenesis; mouse; therapy.
XX
KW
XX
OS Mus musculus.
XX
PN US2003161823-A1.
XX
PD 28-AUG-2003.
XX
PF 14-JAN-2003; 2003US-00341582.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 06-FEB-2001; 2001US-00776874.
PR 05-SEP-2001; 2001WO-IL000830.
PR 19-NOV-2001; 2001US-00988113.
XX
PA (ILAN/) ILAN N.
PA (VLOD/) VLODAVSKY I.
PA (YACO/) YACOBY-ZEEVI O.
PA (PECK/) PECKER I.
PA (FEIN/) FEINSTEIN E.
XX
PI Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;
XX
XX WPI; 2003-897910/82.
DR N-PSDB; ADG88833, ADG88835.
XX
PT Composition for treating a wound comprising recombinant heparanase is
useful to induce or accelerate wound healing and induce or accelerate
angiogenesis.
XX
XX Claim 2; SEQ ID NO 44; 143pp; English.
PS
XX The present invention relates to methods and compositions for inducing
and/or accelerating wound healing via the catalytic activity of
heparanase. The invention is used to induce or accelerate a healing
process, particularly of an ulcer, burn, laceration, surgical incision,
necrosis, pressure wound, diabetic ulcer and to induce or accelerate
angiogenesis. The present sequence is mouse hpa protein.
XX
SQ Sequence 535 AA;
Query Match 100.0%; Score 2797; DB 7; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.9e-262;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLRLLLMLWGPICGALAQGAPAGTAPDDVVDFEYTKRPLRSVSPSFLSTIDASLATD 60
DB 1 MLRLLLMLWGPICGALAQGAPAGTAPDDVVDFEYTKRPLRSVSPSFLSTIDASLATD 60
QY 61 PRFLTFLGSPRLRALARGLPAYLRFGGTKTDFLIFDPDKPTSEERSYMKSQVNHICR 120
DB 61 PRFLTFLGSPRLRALARGLPAYLRFGGTKTDFLIFDPDKPTSEERSYMKSQVNHICR 120
QY 121 SEPVSAVLRKLQVEMFPQELLRLLEQYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFGL 180
DB 121 SEPVSAVLRKLQVEMFPQELLRLLEQYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFGL 180

QY 181 NALLRTTDLRWSSNAQLLLDYCSSKGYNISWELGNPNPSFWKKAHILIDGLQGEDFVE 240
DB 181 NALLRTTDLRWSSNAQLLLDYCSSKGYNISWELGNPNPSFWKKAHILIDGLQGEDFVE 240
QY 241 LHKLLQRSFAQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
DB 241 LHKLLQRSFAQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
QY 301 DFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
DB 301 DFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
QY 361 LGLSAQMGIEVVMQVFPFAGNYHLVDENFPLPDYWLSSLFVKLVGPRVLLSRVKGPDOR 420
DB 361 LGLSAQMGIEVVMQVFPFAGNYHLVDENFPLPDYWLSSLFVKLVGPRVLLSRVKGPDOR 420
QY 421 SKLRVYLHCTNVYHPRYQEGDLTYVLNLHNVTKHLKVPVPLFRKFPVDTYLLKPSGPDGL 480
DB 421 SKLRVYLHCTNVYHPRYQEGDLTYVLNLHNVTKHLKVPVPLFRKFPVDTYLLKPSGPDGL 480
QY 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535
DB 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535
RESULT 4
ADL16413
ID ADL16413 standard; protein; 535 AA.
XX
AC ADL16413;
XX
DT 06-MAY-2004 (first entry)
XX
DE Mouse heparanase protein.
XX
KW Mouse; heparanase; enzyme; heparanase-dependent cancer; cancer;
autoimmune reaction; inflammation.
XX
OS Mus musculus.
XX
PN US2003236215-A1.
XX
PD 25-DEC-2003.
XX
PF 09-JUN-2003; 2003US-00456573.
XX
PR 31-AUG-1998; 98WO-US017954.
PR 01-MAR-1999; 99US-00258892.
PR 08-NOV-1999; 99US-00435739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI; 2004-070610/07.
XX
PT New antisense oligonucleotide hybridizable with a polynucleotide encoding
a polypeptide with heparanase activity, useful for treating diseases such
as cancer and autoimmune disorders.
XX
PS Claim 3; SEQ ID NO 44; 108pp; English.
XX
CC The invention relates to an antisense oligonucleotide (ASO) comprising a
polynucleotide or a polynucleotide analogue of at least 10 bases being
hybridisable in vivo, under physiological conditions, with a portion of
a polynucleotide strand encoding a polypeptide having heparanase
catalytic activity. Also included are a method of in vivo downregulating
heparanase activity (comprising administering the ASO in vivo), a method
of treating a subject suffering from a pathological condition
(characterised by heparanase activity, comprising administering ASO to
the subject), a pharmaceutical composition comprising the ASO and a
carrier, an antisense nucleic acid construct (comprising a promoter

CC sequence and a polynucleotide sequence directing the synthesis of an
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo ,
 CC under physiological conditions, with a polynucleotide strand encoding a
 CC polypeptide having heparanase catalytic activity), a method of in vivo
 CC downregulating heparanase activity (comprising administering in vivo the
 CC antisense nucleic acid construct), a pharmaceutical composition
 CC comprising the antisense nucleic acid construct and a carrier, and an
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide
 CC analogue of at least 10 bases being hybridisable in vivo , under
 CC physiological conditions, with a portion of a polynucleotide strand being
 CC characterised by forming at least a portion of an untranslated region
 CC (UTR) for a polynucleotide strand encoding a polypeptide having
 CC heparanase catalytic activity. The methods and compositions of the
 CC present invention are useful for the prevention and/or treatment of
 CC diseases or conditions associated with aberrant heparanase activity, such
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and
 CC inflammation. The gene for human heparanase is located on chromosome 4.
 CC The present sequence is the mouse heparanase protein.

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGLGALAQAGAPAGTAPDDVDVLEFYTFRPLRSVSPFLSITIDASLATD 60
 DB 1 MLRLLLWLWGLGALAQAGAPAGTAPDDVDVLEFYTFRPLRSVSPFLSITIDASLATD 60

QY 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKSNVNDICR 120
 DB 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKSNVNDICR 120

QY 121 SEPVSAAVLRLKQVEWPFQELLLRQYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180
 DB 121 SEPVSAAVLRLKQVEWPFQELLLRQYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240
 DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLLQSFQNAKLYGPDIGQPRGKTVKLLRSLFKAGGEVIDSLTWHYHLNGRIATKE 300
 DB 241 LHKLLQSFQNAKLYGPDIGQPRGKTVKLLRSLFKAGGEVIDSLTWHYHLNGRIATKE 300

QY 301 DFSSDALDTFILSVQKILKVTKEITPGKKVWLGESAYGGCAPLLSNTFAAGFWLDDK 360
 DB 301 DFSSDALDTFILSVQKILKVTKEITPGKKVWLGESAYGGCAPLLSNTFAAGFWLDDK 360

QY 361 LGLSAQMGLIEVVRQVFFGAGNHYLDENPEPLPDYWLSSLFKLVGPRVLLSRVKGPD 420
 DB 361 LGLSAQMGLIEVVRQVFFGAGNHYLDENPEPLPDYWLSSLFKLVGPRVLLSRVKGPD 420

QY 421 SKRLVYLHCTNVYHPRYQEGDLTYVNLHNVTKHLKVPPLPRKVPDVTYLLKPSGPDGL 480
 DB 421 SKRLVYLHCTNVYHPRYQEGDLTYVNLHNVTKHLKVPPLPRKVPDVTYLLKPSGPDGL 480

QY 481 LSKSVQLNQIILKWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAICI 535
 DB 481 LSKSVQLNQIILKWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAICI 535

RESULT 5

ADM48750

ID ADM48750 standard; protein; 535 AA.

XX

AC ADM48750;

XX

DT 03-JUN-2004 (first entry)

XX

DE Mouse hpa protein.

XX

KW Transgenic animal; heparanase; cancer; viral infection; restenosis;

XX

KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
 KW mouse.

XX Mus musculus.

XX US2003217375-A1.

XX 20-NOV-2003.

XX 24-FEB-2003; 2003US-00371218.

XX 31-AUG-1998; 98WO-US017954.

XX 01-MAR-1999; 99US-00258892.

XX 06-FEB-2001; 2001US-00776874.

XX 19-NOV-2001; 2001US-00986113.

XX (ZCHA/) ZCHARIA E.

XX (VLOD/) VLODAVSKY I.

XX (METZ/) METZGER S.

XX (PECK/) PECKER I.

XX (ILAN/) ILAN N.

XX (CHAJ/) CHAJEK-SHAUL T.

XX (GOLD/) GOLDSHMIDT O.

XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
 Chajek-Shaul T, Goldshmidt O;

XX WPI; 2004-021918/02.

XX N-PSDB; ADM48749, ADM48751.

XX New transgenic non-human animal expressing heparinase, useful as models
 for human disease, such as cancers, viral infection, neurodegenerative
 diseases, restenosis, atherosclerosis and pulmonary disorders.

XX Example 12; SEQ ID NO 44; 106pp; English.

CC The present invention relates to a transgenic non-human animal whose
 genome comprises an exogenous polynucleotide sequence, including a
 CC promoter active in tissues of the non-human, a region encoding a human
 CC heparanase, where the promoter and the region encoding human heparanase
 CC are operably linked in the exogenous polynucleotide such that human
 CC heparanase is expressed in at least a portion of the cells of the non-
 CC human animal. The methods and compositions of the present invention are
 CC useful for the production of transgenic animals expressing heparanase, to
 CC be used as models for human diseases such as cancers, viral infection,
 CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
 CC disorders. The present sequence is mouse hpa protein used in the
 CC exemplification of the invention.

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGLGALAQAGAPAGTAPDDVDVLEFYTFRPLRSVSPFLSITIDASLATD 60
 DB 1 MLRLLLWLWGLGALAQAGAPAGTAPDDVDVLEFYTFRPLRSVSPFLSITIDASLATD 60

QY 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKSNVNDICR 120
 DB 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKSNVNDICR 120

QY 121 SEPVSAAVLRLKQVEWPFQELLLRQYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180
 DB 121 SEPVSAAVLRLKQVEWPFQELLLRQYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240
 DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLLQSFQNAKLYGPDIGQPRGKTVKLLRSLFKAGGEVIDSLTWHYHLNGRIATKE 300
 DB 241 LHKLLQSFQNAKLYGPDIGQPRGKTVKLLRSLFKAGGEVIDSLTWHYHLNGRIATKE 300

Db 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
 QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
 Db 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
 QY 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPDYLWLSLLFKLVGPRVLLSRVKGPD 420
 Db 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPDYLWLSLLFKLVGPRVLLSRVKGPD 420
 QY 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPVPLFRKPVDTYLLKPSGPDGL 480
 Db 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPVPLFRKPVDTYLLKPSGPDGL 480
 QY 481 LSKSVQLNGQILKMWDRQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535
 Db 481 LSKSVQLNGQILKMWDRQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535

RESULT 6

ADR88208
 ID ADR88208 standard; protein; 535 AA.

AC ADR88208;

XX 18-NOV-2004 (first entry)

DT Mouse heparanase.

DE

KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
 KW mouse; heparanase; enzyme.

OS Mus musculus.

XX

Key Location/Qualifiers

FT Peptide

1..17

FT Protein

18..535

FT /label= Signal_peptide

/label= Mature_heparanase

XX

US2004170631-A1.

XX

02-SEP-2004.

XX

28-NOV-2003; 2003US-00722502.

XX

02-SEP-1997; 97US-00922170.

PR

01-MAY-1998; 98US-00071739.

PR

04-NOV-1998; 98US-00186200.

PR

19-FEB-2003; 2003US-00368044.

PR

22-AUG-2003; 2003US-00645659.

XX

(YACO/) YACOBY-ZEEVI O.

PA

(PERE/) PERETZ T.

PA

(MIRO/) MIRON D.

PA

(SHLO/) SHLOMI Y.

PA

(PECK/) PECKER I.

PA

(AYAL/) AYAL-HERSHKOVITZ M.

PA

(FEIN/) FEINSTEIN E.

PA

(VGEL/) VAN GELDER J M.

PA

(VLOD/) VLODAVSKY I.

PA

(FRIE/) FRIEDMANN Y.

XX

Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
 Friedmann Y;

WPI; 2004-625084/60.

XX

Targeted drug delivery to a heparanase-expressing tissue of a patient, useful for treating heparanase-associated conditions such as inflammation or cancer, comprises administering a drug and an anti-heparanase antibody complex.

Claim 2; SEQ ID NO 2; 58pp; English.

The invention relates to a method of targeted drug delivery to a tissue of a patient, the tissue expressing heparanase. The method comprises providing a complex of a drug directly or indirectly linked to an anti-heparanase antibody, and administering the complex to the patient. In the targeted drug delivery, the antibody comprises an epitope of its portion capable of specifically binding to at least one epitope of a heparanase protein. The composition and methods of the invention are useful for diagnosing, preventing or treating conditions associated with heparanase catalytic activity (e.g. an inflammatory disorder, wound, scar, vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell proliferation, invasion of circulating tumour cells and metastatic disease), for purifying heparanase, or for developing drugs for those heparanase-associated conditions. The vasculopathy is atherosclerosis, restenosis or aneurysm. The cancerous condition is a solid cancer or a non-solid cancer. The non-solid cancer is a haematopoietic malignancy selected from acute lymphocytic leukaemia (ALL), acute myelogenous leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia, Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and multiple myeloma. The solid cancer is selected from tumours in lip and oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands, thyroid gland, oesophagus, stomach, small intestine, colon, colorectum, anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva, vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva, malignant melanoma of the conjunctiva, malignant melanoma of the uvea, retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit, brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's sarcoma. The present sequence is mouse heparanase.

Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;

Best Local Similarity 100.0%; Pred. No. 2.9e-262;

Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLGPIGALAQAGAPAGTDDVDLEFYTKPLRSVSPSFLSIITIDASLATD 60
 Db 1 MLRLLLWLGPIGALAQAGAPAGTDDVDLEFYTKPLRSVSPSFLSIITIDASLATD 60
 QY 61 PRFLTFLGSPRLRALRGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDIR 120
 Db 61 PRFLTFLGSPRLRALRGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDIR 120
 QY 121 SEPVSAAVLRKLQVEWPFQELLRLREQYQKFKNSTYSRSSVDMLYSFAKSGLDLIFGL 180
 Db 121 SEPVSAAVLRKLQVEWPFQELLRLREQYQKFKNSTYSRSSVDMLYSFAKSGLDLIFGL 180
 QY 181 NALLRTPDLRNSSNAQLLLDYCSSKGNYSWELGNENPSFWKKAHILIDGLQGEDFVE 240
 Db 181 NALLRTPDLRNSSNAQLLLDYCSSKGNYSWELGNENPSFWKKAHILIDGLQGEDFVE 240
 QY 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
 Db 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
 QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
 Db 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
 QY 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPDYLWLSLLFKLVGPRVLLSRVKGPD 420
 Db 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPDYLWLSLLFKLVGPRVLLSRVKGPD 420

QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLRKRKPYDVTYLLKPSGPDGL 480
Db 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLRKRKPYDVTYLLKPSGPDGL 480
QY 481 LSKSVQLNGQILKMWDEQTLTPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535
Db 481 LSKSVQLNGQILKMWDEQTLTPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535

RESULT 7
ADT78175
ID ADT78175 standard; protein; 535 AA.
XX
AC ADT78175;
DT
DT 13-JAN-2005 (first entry)
XX
DE Mouse heparanase protein.
XX
KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cystostatic; antiinflammatory; vulnery; antiarteriosclerotic;
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; mouse.
XX
OS Mus musculus.

XX Key Location/Qualifiers
FH Binding-site 149..154
FT /note= "Putative heparin binding site"
FT Binding-site 263..269
FT /note= "Putative heparin binding site"
FT Binding-site 418..425
FT /note= "Putative heparin binding site"
XX

US2004213789-A1.
XX
XX 28-OCT-2004.
XX
XX 22-AUG-2003; 2003US-00645659.
XX
XX 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRON/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (GELD/) GELDER J M V.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Herskhovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
XX Claim 5; SEQ ID NO 2; 68pp; English.
XX
XX The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
CC or prostate cancer), minor cell proliferation, invasion of circulating
CC tumour cells, or a metastatic disease, or a heparanase-related disorder
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
CC carcinoma) in a mammal. This sequence represents mouse heparanase.
XX
XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;
Best Local Similarity 100.0%; Pred. No. 2.9e-262;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGLGALAAQAGPAGTPTDDVVDVDFEYTKRPLRSVSPFLSTIDASLATD 60
Db 1 MLRLLLWLWGLGALAAQAGPAGTPTDDVVDVDFEYTKRPLRSVSPFLSTIDASLATD 60
QY 61 PRFLTFLGSPRLALARGLSPAYLRFGGTKTDFLIFDPDKEPTSEERSYKWSQVNHDIR 120
Db 61 PRFLTFLGSPRLALARGLSPAYLRFGGTKTDFLIFDPDKEPTSEERSYKWSQVNHDIR 120
QY 121 SEPVSAAVLRLQVWPFQELLLREQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGL 180
Db 121 SEPVSAAVLRLQVWPFQELLLREQYQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGL 180
QY 181 NALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDEVE 240
Db 181 NALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDEVE 240
QY 241 LHKLQSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
Db 241 LHKLQSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
QY 301 DFLSSDALDFTFILLSVQKILVTKETITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
Db 301 DFLSSDALDFTFILLSVQKILVTKETITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360
QY 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLISLLFKLVGPRVLLSRVKGDPDR 420
Db 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLISLLFKLVGPRVLLSRVKGDPDR 420
QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLRKRKPYDVTYLLKPSGPDGL 480
Db 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLRKRKPYDVTYLLKPSGPDGL 480
QY 481 LSKSVQLNGQILKMWDEQTLTPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535
Db 481 LSKSVQLNGQILKMWDEQTLTPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535

RESULT 8
AEA42424
ID AEA42424 standard; protein; 535 AA.
XX
AC AEA42424;
XX
XX 28-JUL-2005 (first entry)
XX

Mouse heparanase epitope peptide SEQ ID NO:2.

antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;
 antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
 inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
 angiogenesis disorder; cancer; tumor; metastasis.

XX
 OS Mus musculus.

XX AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX XI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of
 PT heparanase protein, useful for preventing and treating heparanase-related
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or
 PT angiogenesis.

XX Claim 2; SEQ ID NO 2; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable
 CC of specifically binding to an epitope of a heparanase protein. Also
 CC described: (1) a cell line (II) for producing a monoclonal antibody or
 CC its portion, comprising a cell line for producing (I); (2) a
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an
 CC affinity medium (III) for binding human heparanase polypeptides,
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
 CC useful for treating a subject suffering from a pathological condition,
 CC which involves administering (I) to the subject. (I) is useful for
 CC preventing and treating heparanase-related disorder or condition chosen
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
 CC cell proliferation, invasion of circulating tumor cells and metastatic
 CC disease. (I) is useful for detecting the presence of heparanase
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related
 CC disease or condition in a subject such as vertebrate, preferably mammal
 CC e.g., human. The heparanase-related disorder or condition further
 CC includes renal disease or disorder chosen from diabetic nephropathy,
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
 CC and renal cell carcinoma. The present sequence represents mouse
 CC heparanase, which is used in the exemplification of the present
 CC invention.

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 9; Length 535;
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLMLWGPGLGALAQAAGTAPTDVDDVLEFYTKRLRSVSPSFLSITDASLATD 60
 DB 1 MLRLLLMLWGPGLGALAQAAGTAPTDVDDVLEFYTKRLRSVSPSFLSITDASLATD 60

QY 61 PRELTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDI 120
 DB 61 PRELTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDI 120

QY 121 SEPVSAVLRLQVWPFQBELLLREYQKEFKNSTYSRSDMLYSFAKSGLDLIFGL 180
 DB 121 SEPVSAVLRLQVWPFQBELLLREYQKEFKNSTYSRSDMLYSFAKSGLDLIFGL 180

QY 181 NALLRTPDLRNWNSSNAQLLLDYCSSKGYNIWELGNEPNSFWKKAHILIDGLQGEDFVE 240
 DB 181 NALLRTPDLRNWNSSNAQLLLDYCSSKGYNIWELGNEPNSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLQSAFQNAKLYGPDIGQPRGKTVKLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
 DB 241 LHKLQSAFQNAKLYGPDIGQPRGKTVKLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300

QY 301 DFLSSDALDTPILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWLDDK 360
 DB 301 DFLSSDALDTPILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWLDDK 360

QY 361 LGLSAQMGIEVVMROVFFGAGNYHLVDENFPLPDYWLSSLFVKLVGRVLLSRVKGPDOR 420
 DB 361 LGLSAQMGIEVVMROVFFGAGNYHLVDENFPLPDYWLSSLFVKLVGRVLLSRVKGPDOR 420

QY 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPPLFRKPPVDTYLLKPSGPDGL 480
 DB 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPPLFRKPPVDTYLLKPSGPDGL 480

QY 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535
 DB 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535

RESULT 9
 ADY27033
 ID ADY27033 standard; protein; 535 AA.

XX AC ADY27033;

XX DT 05-MAY-2005 (first entry)

XX DE Murine heparanase protein.

XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.

XX OS Mus musculus.

XX PN WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase
 PT neurological diseases or viral diseases) comprises modulating heparanase
 PT activation.

XX Disclosure; SEQ ID NO 5; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a
 CC tissue or regulating a biological process depending at least in part on
 CC heparanase activity comprising modulating heparanase activation. The
 CC invention also relates to methods of treating a heparanase- or heparin
 CC binding protein-associated disease or disorder in a subject, a
 CC pharmaceutical composition for use in the treatment of a heparanase-
 CC associated disease or disorder comprising a therapeutic amount of an
 CC agent capable of modulating heparanase activation and a pharmaceutical

CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a murine heparanase protein used in the scope of the
CC invention.
XX
XX Sequence 535 AA;
SQ

Query Match 99.9%; Score 2793; DB 9; Length 535;
Best Local Similarity 99.8%; Pred. No. 7.2e-262;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGPGLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLATD 60
DB 1 MLRLLLWLWGPGLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLATD 60

QY 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTTKDFLI FDPDKPTSEERSYKWSQVNHDI 120
DB 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTTKDFLI FDPDKPTSEERSYKWSQVNHDI 120

QY 121 SEPVSAAVLRLQVWPFQFELLRLRQYQKEFNKSTYSRSSVDMLYSFAKCSGLDLIFGL 180
DB 121 SEPVSAAVLRLQVWPFQFELLRLRQYQKEFNKSTYSRSSVDMLYSFAKCSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240
DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300
DB 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300

QY 301 DFLSSDALDTFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPILLSNTFAAGFMWLDK 360
DB 301 DFLSSDVLTFTILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPILLSNTFAAGFMWLDK 360

QY 361 LGLSAQMGIQEVNMRQVFFGAGNYHLVDENFPELPDYWLSLLFKKLVGRVLLSRVKGPD 420
DB 361 LGLSAQMGIQEVNMRQVFFGAGNYHLVDENFPELPDYWLSLLFKKLVGRVLLSRVKGPD 420

QY 421 SKLRVYLHCTNVHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDGL 480
DB 421 SKLRVYLHCTNVHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDGL 480

QY 481 LSKSVOLNGOILKMVDEQTLPALTEKPLPAGSALSIPAFSGYGFVIRNAKIAACI 535
DB 481 LSKSVOLNGOILKMVDEQTLPALTEKPLPAGSALSIPAFSGYGFVIRNAKIAACI 535

RESULT 10
ABB07812
ID ABB07812 standard; protein; 536 AA.
XX
AC ABB07812;
XX
DT 03-JUL-2002 (first entry)
XX
DE Rat heparanase sequence.
XX
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; rat.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT Peptide 1..16
FT /note= "putative signal peptide"

FT Protein 17..536
FT /note= "mature protein"
XX US2002034810-A1.
XX 21-MAR-2002.
XX
XX 16-AUG-2001; 2001US-00930218.
XX 20-SEP-2000; 2000US-00666390.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;
XX WPI; 2002-338926/37.
XX
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
XX to treat various heparin-related disorders and the signal peptide is
XX useful in production of membrane-targeted or secreted recombinant
XX proteins.
XX
XX Disclosure; Fig 1a; 39pp; English.
XX
XX The invention relates to an isolated avian and reptile nucleic acid,
XX encoding a polypeptide with heparanase catalytic activity. The signal
XX peptide of the nucleic acid can be used to express membrane-associated or
XX secreted proteins in heterologous expression systems. The encoded
XX polypeptides can be used to prevent tumour angiogenesis, metastasis and
XX invasion, and to intervene with pathologies associated with impaired
XX heparin-binding growth factors, cellular responses to heparin-binding
XX growth factors and cytokines, cell interaction with plasma lipoproteins,
XX cellular susceptibility to viral, protozoa and bacterial infections or
XX disintegration of neurodegenerative plaques. The present sequence
XX represents a rat heparanase protein sequence used in similarity studies
XX
XX Sequence 536 AA;
SQ

Query Match 92.6%; Score 2590.5; DB 5; Length 536;
Best Local Similarity 92.5%; Pred. No. 3.5e-242;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 1 MLR-LLLLWLWGPGLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59
DB 1 MLRPLLLLLWLWGRALRALTQGTAGTAPTKDQVLDLEFYTKRPLRSVSPFLSITIDASLAT 60

QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTTKDFLI FDPDKPTSEERSYKWSQVNHDI 119
DB 61 DPRFLTFLGSPRLRALARGLSPAYLRFGGTTKDFLI FDPDKPTSEERSYKWSQVNHDI 120

QY 120 RSEPVSAAVLRKLQVWPFQFELLRLRQYQKEFNKSTYSRSSVDMLYSFAKCSGLDLIFG 179
DB 121 GSERVSAVLRKLQVWPFQFELLRLRQYQKEFNKSTYSRSSVDMLYSFAKCSGLDLIFG 180

QY 180 INALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239
DB 181 LNALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240

QY 240 ELHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATK 299
DB 241 ELHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATK 300

QY 300 EDFLSSDALDTFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPILLSNTFAAGFMWLD 359
DB 301 EDFLSSDVLTFTILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPILLSNTFAAGFMWLD 360

QY 360 KGLSAQMGIQEVNMRQVFFGAGNYHLVDENFPELPDYWLSLLFKKLVGRVLLSRVKGPD 419
DB 361 KGLSAQMGIQEVNMRQVFFGAGNYHLVDENFPELPDYWLSLLFKKLVGRVLLSRVKGPD 420

QY 420 RSKLRVYLHCTNVHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDG 479
DB 421 RSKLRVYLHCTNVHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDG 480

[illegible]

CC	protein. The composition and methods of the invention are useful for
CC	diagnosing, preventing or treating conditions associated with heparanase
CC	catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC	vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC	proliferation, invasion of circulating tumour cells and metastatic
CC	disease), for purifying heparanase, or for developing drugs for those
CC	heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC	restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC	non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC	selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC	leukemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC	leukemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC	Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC	multiple myeloma. The solid cancer is selected from tumours in lip and
CC	oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC	thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC	anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC	Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC	sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC	vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC	tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC	bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC	malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC	retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC	brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC	sarcoma. The present sequence is rat heparanase.
XX	
SQ	Sequence 536 AA;
	Query Match 92.6%; Score 2590.5; DB 8; Length 536;
	Best Local Similarity 92.5%; Pred. No. 3.5e-242;
	Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;
QY	1 MLR-LLLLWLWGLGALAAQAGTAGTADTDVVDLEFYTKRPLRSVSFSLITIDASLAT 59
DB	1 MLRPLLLLLWGLRLALTQTGTAGTAPTKDQVDLEFYTKRLFQSVSFSLITIDASLAT 60
QY	60 DPRFLTLFGSPRLRALARGLSPAYLRPGGYTKDTFLIPDPKPTSEERSYWKSVQHNDIC 119
DB	61 DPRFLTFLGSPRLRALARGLSPAYLRPGGYTKDTFLIPDNKEPTSEERSYWKSDNNDIC 120
QY	120 RSPVSAVLRLQVEMPFQELLRLREOYKFKNYSRSSVDMLYSFAKCSGLDLIFG 179
DB	121 GSRVSADVLRKLQWEMPFQELLRLREOYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180
QY	180 LNALLRTPDLRWNSSNAQLLDLYCSSKGYNISWELGNPNPSFWKAHILIDGLQLGEDFV 239
DB	181 LNALLRTPDLRWNSSNAQLLLNYCSSKGYNISWELGNPNPSFWKKAQISIDGLQLGEDFV 240
QY	240 ELHKLLQRSFAFNAKLYGPDIGQPRGTVTKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299
DB	241 ELHKLLQKSFAFNAKLYGPDIGQPRGTVTKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
QY	300 EDFLSSDALDTFILSVOKILKVTKETTPGKKVWLGETSSAYGGCAPILLSNTFAAGFWMLD 359
DB	301 EDFLSSDVLDFTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGCAPULSNTFAAGFWMLD 360
QY	360 KLGLSQMGIEVVMRQVFVGAGNYHLVDENFEPLPDYWLSSLFKKLGVPRVLLSRVRKGD 419
DB	361 KLGLSQAQLGIEVVMRQVFVGAGNYHLVDENFEPLPDYWLSSLFKKLGVPKVLMRSVRKGD 420
QY	420 RSKLRVYLHCTNVYHPRYQBGDLTLVYVLNLHNVTYKHLKVPPDPIFRKPVDYTYLLKPSGPDG 479
DB	421 RSKLRVYLHCTNVYHPRYREGDLTLVYVLNLHNVTYKHLKLPPPMFSRFPDYLYLLKPFSGDG 480
QY	480 LLSKSVQLNGQILLKWDEQTLPALTEKLPAGSALSIPAESYGFFVIRNKAIAACI 535
DB	481 LLSKSVQLNGQTLIKWDEQTLPALTEKLPAGSSLSVPAFSYGFFVIRNKAIAACI 536
RESULT 12	
ADT781176	
ID ADT781176 standard; protein; 536 AA.	

XX ADT78176;
 XX 13-JAN-2005 (first entry)
 XX Rat heparanase protein.
 XX Antibody; epitope; heparanase; pathological condition; angiogenesis;
 KW cell proliferation; cancerous condition; tumour cell invasion;
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;
 KW cystostatic; antiinflammatory; vulnerable; antiarteriosclerotic;
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; rat.
 XX Rattus norvegicus.
 XX Key Location/Qualifiers
 FH Binding-site 150..155
 FT /note= "Putative heparin binding site"
 FT Binding-site 264..270
 FT /note= "Putative heparin binding site"
 FT Binding-site 419..426
 FT /note= "Putative heparin binding site"
 XX US2004213789-A1.
 PN 28-OCT-2004.
 XX 22-AUG-2003; 2003US-00645659.
 XX 02-SEP-1997; 97US-00922170.
 PR 01-MAY-1998; 98US-00071739.
 PR 04-NOV-1998; 98US-00186200.
 PR 19-FEB-2003; 2003US-00368044.
 XX (YACO/) YACOBY-ZEVI O.
 PA (PERE/) PERETZ T.
 PA (MIRO/) MIRON D.
 PA (SHLO/) SHLOMI Y.
 PA (PECK/) PECKER I.
 PA (AYAL/) AYAL-HERSHKOVITZ M.
 PA (FEIN/) FEINSTEIN E.
 PA (GELD/) GELDER J M V.
 PA (VLOD/) VLODAVSKY I.
 PA (FRIE/) FRIEDMANN Y.
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodaysky I;
 PI Friedmann Y;
 XX WPI; 2004-774790/76.
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for
 PT detecting, treating or preventing cancer, inflammatory or autoimmune
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
 XX Claim 5; SEQ ID NO 3; 68pp; English.
 XX The invention relates to an isolated antibody or antibody portion capable
 CC of specifically binding to or elicited by at least one epitope of a
 CC heparanase protein, where the heparanase protein is at least 60%
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
 CC where at least one epitope comprises a sequence of at least 70% homologous
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
 CC a hybridoma cell line comprising a cell line for producing the monoclonal
 CC antibody, (b) a method for detecting, treating or preventing a
 CC pathological condition or a heparanase-related disorder or condition in a
 CC subject, (c) a method for monitoring the state of a heparanase-related
 CC disorder or condition in a subject, and (d) a pharmaceutical composition
 CC comprising the isolated anti-heparanase antibody or antibody portion and
 CC a pharmaceutical carrier. The antibody, methods, and composition are
 CC useful for detecting, treating, preventing or monitoring a pathological
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents rat heparanase.
 XX
 SQ Sequence 536 AA;
 Query Match 92.6%; Score 2590.5; DB 8; Length 536;
 Best Local Similarity 92.5%; Pred. No. 3.5e-242;
 Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;
 QY 1 MLR-LLLMLWGPGLGALAQAPAGTAPTDVVDLEFYTCKRPLRSVSPFLSITIDASLAT 59
 DB 1 MLRPLLLLLWGLRLALTQGTAGTAPTKDQVDLEFYTCKLFQSVSPFLSITIDASLAT 60
 QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFEGGTTKDFLFDPPKPTSEERSYWKSVQNHQIC 119
 DB 61 DPRFLTFLGSPRLRALARGLSPAYLRFEGGTTKDFLFDPNKEPTSEERSYWQSQNDNIC 120
 QY 120 RSEPVSAVLRLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYSFAKCSRLDLIFG 179
 DB 121 GSERVSADVLRLQVWPFPQELLRLREYQKQKFNSTYSRSSVDMLYSFAKCSRLDLIFG 180
 QY 180 LNALLRTPDLRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
 DB 181 LNALLRTPDLRWNSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
 QY 240 ELHKLLQRSFAONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
 DB 241 ELHKLLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRVATK 300
 QY 300 EDFLSSDALDFTFILSVQKILKVTKEITPGKKVWLGGETSSAYGGGAPLLSNTFAAGFMWLD 359
 DB 301 EDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLGGETSSAYGGGAPLLSNTFAAGFMWLD 360
 QY 360 KLGLSAQMGIEVVMQVFFGAGNVLVDENFELPDYWLNLHNTVKHLKVPPLPRKVPDVTLLKPSGPD 419
 DB 361 KLGLSAQMGIEVVMQVFFGAGNVLVDENFELPDYWLNLHNTVKHLKVPPLPRKVPDVTLLKPSGPD 420
 QY 420 RSKLRVYLHCTNVYHPRYQEGDLTYVLNLHNTVKHLKVPPLPRKVPDVTLLKPSGPD 479
 DB 421 RSKLRVYLHCTNVYHPRYQEGDLTYVLNLHNTVKHLKVPPLPRKVPDVTLLKPSGPD 480
 QY 480 LLSKSVQLNGQILKMVDEQTLPALETEKPLPAGSALSPLAFSYGFFVIRNAKIAACI 535
 DB 481 LLSKSVQLNGQILKMVDEQTLPALETEKPLPAGSALSPLAFSYGFFVIRNAKIAACI 536
 RESULT 13
 ADY27035
 ID ADY27035 standard; protein; 536 AA.
 XX AC ADY27035;
 XX DT 05-MAY-2005 (first entry)
 XX DE Rat heparanase protein.
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
 KW neurological disease; viral infection; infection; cytostatic;
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
 KW protease; enzyme; enzyme purification.
 XX Rattus norvegicus.
 XX WO2005016227-A2.
 XX 24-FEB-2005.

PF	12-AUG-2004; 2004WO-IL000744.
XX	
PR	14-AUG-2003; 2003US-0494800P.
PR	12-JAN-2004; 2004US-0535492P.
XX	
PA	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX	
PI	Van-Gelder JM, Miron D;
XX	
DR	WPI; 2005-182203/19.
XX	
XX	Regulating heparanase activity, useful for treating heparanase-associated diseases (e.g. cancer, inflammation, cardiovascular diseases, neurological diseases or viral diseases) comprises modulating heparanase activation.
PT	
PT	
XX	
PS	Disclosure; SEQ ID NO 7; 211pp; English.
XX	
CC	The invention relates to a method of regulating heparanase activity in a tissue or regulating a biological process depending at least in part on heparanase activity comprising modulating heparanase activation. The invention also relates to methods of treating a heparanase- or heparin binding protein-associated disease or disorder in a subject, a pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence represents a rat heparanase protein used in the scope of the invention.
XX	
SQ	Sequence 536 AA;
	Query Match 92.6%; Score 2590.5; DB 9; Length 536;
	Best Local Similarity 92.5%; Pred. No. 3.5e-242;
	Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1
Qy	1 MLR-LLLLLWGPGALAAQAAGTAPTDVVLDLEFYTKRLRSVSPSLITIDASLAT 59
Dd	1 MLRP LLLLWLWGRLLALTQTGTPACTPTKOVVDLEFYTKRLFQSVSPSLITIDASLAT 60
Qy	60 DPRPLTFGLSGPRLRALAGLSPAYLRFGTGKTDFLIPDDKEPTSEERSYKVSQNHDIC 119
Dd	61 DPRPLTFGLSGPRLRALAGLSPAYLRFGTGKTDFLIPDNKEPTSEERSYKQSNDNIC 120
Qy	120 RSPVSAVLRLKLOVEVPFQELLRLRQYQKEFNKNSTYSRSSVDMLYSFACSGLDLIFG 179
Dd	121 GSERVSAVLRLKLQWEVPFQELLLRLRQYQREFKNSTYSRSSVDMLYSFACSRDLIFG 180
Qy	180 LNALLRTPDLRWNSSNAQLLDYCYSKGYNISWELGNPNFWFKKAHLIIDGLQAGEDVF 239
Dd	181 LNALLRTPDLRWNSSNAQLLLNYCCKGYNISWELGNPNFWFKKAQSIDGLQGEDVF 240
Qy	240 ELHKLLQRSFAONAKLYGPDIGQRGKTVKLLRSFLKAGGEVIDSLTWHHYYLNGRIATK 299
Dd	241 ELHKLLQKSFAONAKLYGPDIGQRGKTVKLLRSFLKAGGEVIDSLTWHHYYLNGRVATK 300
Qy	300 EDFLSSDALDTFILSVOKILKVTEITPGKKVMJGETSSAYGGGAPILLSNTFAAGFMWLD 359
Dd	301 EDFUSSDVLDTFILSVOKILKVTEMTPGKKVMJGETSSAYGGGAPILLSNTFAAGFMWLD 360
Qy	360 KLGLSAQMGIENVNRQVFFGAGNTHLVDENPEPLPDYWLSSLFLFKLVGPRLVRKVGPD 419
Dd	361 KLGLSAQGLGIEVNRQVFFGAGNTHLVDENPEPLPDYWLSSLFLFKLVGPRLVSRVGPD 420
Qy	420 RSKLRVYLHCTNVTHRYQEGLDTLYVLNLHNVTGHUKVPPPLFRKPVDVTVLLKPSGDG 479

421	RSKURVYLHCTNNVHPRYREGD	LTLYVLNLNHNVTXKHLKLP	PPMFRSPVDKYLLKPGSDG	480
480	LLSKSVQLNGQILKQWVDEQ	TLPALTEKPLPAGSALS	PAFSYGVFFVIRNAKIAACI	535
481	LLSKSVQLNGQTLKQWVDEQ	TLPALTEKPLPAGSSLS	VPFAFSYGVFFVIRNAKIAACI	536
<p>RESULT 14</p> <p>AEA42425</p> <p>ID AEA42425 standard; protein; 536 AA.</p> <p>AC AEA42425;</p> <p>XX 28-JUL-2005 (first entry)</p> <p>XX Rat heparanase epitope peptide SEQ ID NO:3.</p> <p>DE</p> <p>XX antibody; heparanase; antiinflammatory; vulnerary; immunosuppressive;</p> <p>XX antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;</p> <p>KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;</p> <p>KW angiogenesis disorder; cancer; tumor; metastasis.</p> <p>XX</p> <p>OS Rattus norvegicus.</p> <p>XX</p> <p>XX AU2004201462-A1.</p> <p>XX</p> <p>XX 06-MAY-2004.</p> <p>XX</p> <p>XX 08-APR-2004; 2004AU-00201462.</p> <p>PF</p> <p>XX 08-APR-2004; 2004AU-00201462.</p> <p>XX</p> <p>XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.</p> <p>PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.</p> <p>XX</p> <p>XX Vladavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;</p> <p>PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;</p> <p>PI Feinstein E;</p> <p>XX</p> <p>DR WPI; 2005-173343/19.</p> <p>XX</p> <p>XX Novel isolated antibody capable of specifically binding to epitope of</p> <p>PT heparanase protein, useful for preventing and treating heparanase-related</p> <p>PT disorder such as inflammatory disorder, scars, autoimmune conditions or</p> <p>PT angiogenesis.</p> <p>XX</p> <p>PS Claim 2; SEQ ID NO 3; 260pp; English.</p> <p>XX</p> <p>XX The invention relates to an isolated antibody or its portion (I) capable</p> <p>CC of specifically binding to an epitope of a heparanase protein. Also</p> <p>CC described: (1) a cell line (II) for producing a monoclonal antibody or</p> <p>CC its portion, comprising a cell line for producing (I); (2) a</p> <p>CC pharmaceutical composition comprising (I) and a carrier; and (3) an</p> <p>CC affinity medium (III) for binding human heparanase polypeptides,</p> <p>CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)</p> <p>CC useful for treating a subject suffering from a pathological condition,</p> <p>CC which involves administering (I) to the subject. (I) is useful for</p> <p>CC preventing and treating heparanase-related disorder or condition chosen</p> <p>CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune</p> <p>CC condition, angiogenesis, cell proliferation, cancerous condition, tumor</p> <p>CC cell proliferation, invasion of circulating tumor cells and metastatic</p> <p>CC disease. (I) is useful for detecting the presence of heparanase</p> <p>CC polypeptide in a sample. (I) is useful for detecting heparanase-related</p> <p>CC disease or condition in a subject such as vertebrate, preferably mammal</p> <p>CC e.g., human. The heparanase-related disorder or condition further</p> <p>CC includes renal disease or disorder chosen from diabetic nephropathy,</p> <p>CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome,</p> <p>CC and renal cell carcinoma. The present sequence represents rat heparanase,</p> <p>CC which is used in the exemplification of the present invention.</p> <p>XX</p> <p>SQ Sequence 536 AA;</p>				

Query Match 92.6%; Score 2590.5; DB 9; Length 536;

Best Local Similarity 92.5%; Pred. No. 3.5e-242;		Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	
QY	1	MLR-LLLLMWGLGALAAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT	59
DB	1	MLRPLLLMLWGLRALTQGTAGTAPTKDQVLEFYTKRPLRSVSPSFLSITIDASLAT	60
QY	60	DRFLFLGSPRLALARGSPAYLRFGGTTKDFLIPDPKEPTSEERSYKQVNHDC	119
DB	61	DRFLFLGSPRLALARGSPAYLRFGGTTKDFLIPDPKEPTSEERSYKQVNHDC	120
QY	120	RSPVSAALVRKLQVWPFQELLRLREYQKFNSTYSRSDVMLYSPAKCSGLDLIFG	179
DB	121	GSRVSADVLRKLQVWPFQELLRLREYQKFNSTYSRSDVMLYSPAKCSGLDLIFG	180
QY	180	LNALLTPDLRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	239
DB	181	LNALLTPDLRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	240
QY	240	ELHKLQSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK	299
DB	241	ELHKLQSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK	300
QY	300	EDFLSSDALDTFTLSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD	359
DB	301	EDFLSSDVLDFTFTLSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD	360
QY	360	KLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRVLLSRVKGPD	419
DB	361	KLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRVLLSRVKGPD	420
QY	420	RSKLRVYLHCTNVYHPRYQEGDITLYVLNLHNTVTKHLKVPPLFRKPVDTYLLKPSGPDG	479
DB	421	RSKLRVYLHCTNVYHPRYQEGDITLYVLNLHNTVTKHLKVPPLFRKPVDTYLLKPSGPDG	480
QY	480	LLSKSVQLNGQILKWVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI	535
DB	481	LLSKSVQLNGQILKWVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI	536
RESULT 15			
ID	ADY27034	standard; protein; 545 AA.	
XX	ADY27034;		
AC	ADY27034;		
XX	ADY27034;		
DT	05-MAY-2005 (first entry)		
XX	Bovine heparanase protein.		
XX	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;		
KW	neurological disease; viral infection; infection; cytostatic;		
KW	antiflammatory; cardiovascular-gen.; neuroprotective; virucide;		
XX	protease; enzyme; enzyme purification.		
OS	Bos taurus.		
XX	WO2005016227-A2.		
PN	24-FEB-2005.		
XX	12-AUG-2004; 2004WO-IL000744.		
XX	14-AUG-2003; 2003US-0494800P.		
PR	12-JAN-2004; 2004US-0535492P.		
XX	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.		
PA	Van-Gelder JM, Miron D;		
XX	WPI; 2005-182203/19.		
XX	Regulating heparanase activity, useful for treating heparanase-associated		

PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,		
PT	neurological diseases or viral diseases) comprises modulating heparanase		
PT	activation.		
PS	Disclosure; SEQ ID NO 6; 211pp; English.		
XX	The invention relates to a method of regulating heparanase activity in a		
CC	tissue or regulating a biological process depending at least in part on		
CC	heparanase activity comprising modulating heparanase activation. The		
CC	invention also relates to methods of treating a heparanase- or heparin		
CC	binding protein-associated disease or disorder in a subject, a		
CC	pharmaceutical composition for use in the treatment of a heparanase-		
CC	associated disease or disorder comprising a therapeutic amount of an		
CC	agent capable of modulating heparanase activation and a pharmaceutical		
CC	carrier or diluent, a method of identifying a protease activator of		
CC	heparanase, a protease substrate mimetic comprising a peptide		
CC	representing a subset or all substrate residues or cleavage sites of		
CC	human heparanase or an equivalent non-human heparanase, a method of		
CC	producing active heparanase and a method of modulating an adhesion		
CC	activity of heparanase. The composition and methods are useful for		
CC	modulating heparanase activation and for treating heparanase-associated		
CC	diseases or disorders such as cancer, inflammation, cardiovascular		
CC	diseases, neurological diseases or viral infections. This sequence		
CC	represents a bovine heparanase protein used in the scope of the		
CC	invention.		
XX	Sequence 545 AA;		
SQ	Query Match 78.1%; Score 2184; DB 9; Length 545;		
	Best Local Similarity 77.8%; Pred. No. 1.2e-202;		
	Matches 414; Conservative 46; Mismatches 72; Indels 0; Gaps 0;		
QY	4	LLLLLWGLGALAAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLATDRF	63
DB	14	LLLLLWGLGALAAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLATDRF	73
QY	64	LTFGLSPRLALARGSPAYLRFGGTTKDFLIPDPKEPTSEERSYKQVNHDCRSE	123
DB	74	FTFLGSSKLTARGLAPAYLRFGGTTKDFLIPDPKEPTSEERSYKQVNHDCRSE	133
QY	124	VSAVLRKLQVWPFQELLRLREYQKFNSTYSRSDVMLYSPAKCSGLDLIFGNAL	183
DB	134	IPSDVEEKLSEWPFQELLRLREYQKFNSTYSRSDVMLYSPAKCSGLDLIFGNAL	193
QY	184	LRTPLDRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	243
DB	194	LRTDMHWSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	253
QY	244	LLQSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLNGRIATKEDFL	303
DB	254	LLKSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLNGRIATKEDFL	313
QY	304	SSDALDTFTLSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD	363
DB	314	NPDLDTFTLSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD	373
QY	364	SAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRVLLSRVKGPD	423
DB	374	SARMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRVLLSRVKGPD	433
QY	424	RVYLHCTNVYHPRYQEGDITLYVLNLHNTVTKHLKVPPLFRKPVDTYLLKPSGPD	483
DB	434	RVYLHCTNVYHPRYQEGDITLYVLNLHNTVTKHLKVPPLFRKPVDTYLLKPSGPD	493
QY	484	SVQLNGQILKWVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI	535
DB	494	SVQLNGQILKWVDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI	545
Search completed: June 5, 2006, 12:09:44			
Job time : 107.661 secs			

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 95.7733 Seconds
(without alignments)
3728.138 Million cell updates/sec

Title: US-10-645-659A-1
Perfect score: 2020
Sequence: 1 KFKFNSTYSRSSVDVLYTFA.....LPAFSYSFVIRNAKVAACI 386

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2020	100.0	543	1	HPSE HUMAN
2	1724	85.3	574	2	Q333X9 9RODE
3	1723	85.3	574	2	Q333X8 9RODE
4	1711	84.7	574	2	Q333X7 9RODE
5	1710	84.7	574	2	Q333X6 9RODE
6	1681	83.2	545	1	HPSE BOVIN
7	1661	82.2	535	1	HPSE MOUSE
8	1646	81.5	536	1	HPSE RAT
9	1602	79.3	558	2	Q333X5 SPAJD
10	1345.5	66.6	523	1	HPSE CHICK
11	1059.5	52.5	533	2	Q4SIF6 TETNG
12	978	48.4	592	1	HPSE2 HUMAN
13	978	48.4	592	2	Q2MLH9 HUMAN
14	874	43.3	597	2	Q4TB80 TETNG
15	742.5	36.8	255	2	Q4TC8 TETNG
16	622.5	30.8	515	2	Q8TI08 BOMMO
17	407	20.1	543	1	HPSE1 ARATH
18	379.5	18.8	526	2	Q5SNA6 ORYSA
19	354.5	17.5	536	1	HPSE3 ARATH
20	354	17.5	401	2	Q30324 ARATH
21	352	17.4	559	2	Q89F99 BRAJA
22	349.5	17.3	541	2	Q69116 ORYSA
23	349	17.3	539	2	Q2QN56 ORYSA
24	347.5	17.2	537	2	Q70YJ3 HORVU
25	345.5	17.1	539	1	HPSE2 ARATH
26	345	17.1	529	2	Q6ZJE2 ORYSA
27	339	16.8	527	2	Q9LRC8 SCUBA
28	290	14.4	516	2	Q447R5 SOLUS
29	237.5	11.8	506	2	Q37Q70 SPHAR
30	237	11.7	382	2	Q3E8P7 ARATH
31	236.5	11.7	537	2	Q43S03 SOLUS

32	156	7.7	935	2	Q9VE79 DROME	Q9VE79 drosophila
33	135	6.7	1128	2	Q5T65 ANOGA	Q5T65 anopheles g
34	130	6.4	463	2	Q3T97 BURPS	Q3T97 burkholderi
35	130	6.4	670	2	Q3JTG0 BURP1	Q3JTG0 burkholderi
36	126.5	6.3	493	2	Q9HK01 THEAC	Q9HK01 thermoplas
37	124	6.1	510	2	Q2U0T3 ASPOR	Q2U0T3 aspergillus
38	123	6.1	634	2	Q5N1J7 9BACT	Q5N1J7 uncultured
39	122	6.0	559	2	Q7SP80 NEUCR	Q7SP80 neurospora
40	119	5.9	815	2	Q4GJ2 DICDI	Q4GJ2 dictyosteli
41	116	5.7	390	2	Q8TPH7 METAC	Q8TPH7 methanosarc
42	112.5	5.6	489	2	Q5DYH2 VIBF1	Q5DYH2 vibrio fisc
43	111	5.5	536	2	Q2UDS9 ASPOR	Q2UDS9 aspergillus
44	110.5	5.5	653	2	Q3NZS7 9GAMM	Q3NZS7 shewanella
45	110	5.4	665	2	Q5S1C3 CRINE	Q5S1C3 cryptococcu

ALIGNMENTS

RESULT 1
HPSE HUMAN
ID HPSE HUMAN STANDARD; PRT; 543 AA.
AC Q9Y251; Q53GE5; Q9UL39;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1999, sequence version 1.
DT 07-FEB-2006, entry version 27.
DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpa1) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
GN Name=HPSE; Synonyms=HEP, HPA, HPAL, HPRI, HPSE1, HSE1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;
RT "Cloning and functional expression of a human heparanase gene.";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
[2]
RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.
RP SEQUENCE OF 158-168; 326-337 AND 447-491.
RC TISSUE=Embryonic fibroblast;
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and expression.";
RL J. Biol. Chem. 274:24153-24160(1999).
[3]
RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.
RX PubMed=10395325; DOI=10.1038/10518;
RA Vlodaysky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Iehai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Becker I.;
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";
RL Nat. Med. 5:793-802(1999).
[4]
RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";
RL Nat. Med. 5:803-809(1999).
[5]
RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE

RP TISSUE=Placenta;
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
RT "Heparanase expression in invasive trophoblasts and acute vascular
damage.";
RL Glycobiology. 10:467-475(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
RX PubMed=11547900; DOI=10.1023/A:1011375624902;
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
RA Aviv A., Ekin M., Pecker I., Peretz T., Vlodavsky I.;
RT "Molecular properties and involvement of heparanase in cancer
progression and mammary gland morphogenesis.";
RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).
RN [7]
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
RC TISSUE=Placenta;
RX PubMed=12713442; DOI=10.1042/BJ20030318;
RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,
RA Snell D., Schofield C., Gong H., Townsend R., Gallagher J., Page M.,
RA Parakh R., Stubberfield C.;
RT "Biochemical characterization of the active heterodimer form of human
heparanase (HpaI) protein expressed in insect cells.";
RL Biochem. J. 373:423-435(2003).
RN [8]
RP NUCLEOTIDE SEQUENCE [MRNA].
RA Pinhal M.A., Smedo P.;
RT "Cloned heparanase from MCF-7 cells.";
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
RN [9]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Small intestine;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
RN [10]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Rutherford Y.S.N., Krzywinski M.I., Skalek U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [11]
RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
RX PubMed=11123890; DOI=10.1021/bi002080p;
RA Hullett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
RA Gready J.E., Parish C.R.;
RT "Identification of active-site residues of the pro-metastatic
endoglycosidase heparanase.";
RN Biochemistry 39:15659-15667(2000).
RN [12]
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
RP ASN-217; ASN-238 AND ASN-459.

RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
RA Smizu S., Ishida K., Wierzb M.K., Osada H.;
RT "Secretion of heparanase protein is regulated by glycosylation in
human tumor cell lines.";
RL J. Biol. Chem. 279:2697-2703(2004).
RN [13]
RP SUBCELLULAR LOCATION.
RX PubMed=15292202; DOI=10.1074/jbc.M402131200;
RA Gangis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
RT "Heparanase uptake is mediated by cell membrane heparan sulfate
proteoglycans.";
RL J. Biol. Chem. 279:44084-44092(2004).
RN [14]
RP BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
RX PubMed=1584168; DOI=10.1016/j.febslet.2005.03.030;
RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
RT "Heparanase processing by lysosomal/endosomal protein preparation.";
RL FEBS Lett. 579:2334-2338(2005).
RN [15]
RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;
RA Abboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
RA Elgavish S., Peretz T., Vlodavsky I.;
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
human proheparanase.";
RL J. Biol. Chem. 280:13568-13575(2005).
RN [16]
RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;
RA Levy-Adam F., Abboud-Jarrous G., Guerrini M., Beccati D.,
RA Vlodavsky I., Ilan N.;
RT "Identification and characterization of heparin/heparan sulfate
binding domains of the endoglycosidase heparanase.";
RL J. Biol. Chem. 280:20457-20466(2005).
RN [17]
RP VARIANT SER-260.
RX PubMed=15334672;
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
RA Shi L.H., Shen H.W., Xu B.;
RT "Heparanase mRNA expression and point mutation in hepatocellular
carcinoma.";
RL World J. Gastroenterol. 10:2795-2799(2004).
RN [18]
CC -I- FUNCTION: Endoglycosidase which is a cell surface and
extracellular matrix-degrading enzyme. Cleaves heparan sulfate
proteoglycans (HSPGs) into heparan sulfate side chains and core
proteoglycans. Also implicated in the extravasation of leukocytes
and tumor cell lines. Due to its contribution to metastasis and
angiogenesis, it is considered to be a potential target for anti-
cancer therapies.
CC -I- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
lower extent, by heparin and sulfamin and activated by calcium and
magnesium (By similarity).
CC -I- BIOPHYSICOCHEMICAL PROPERTIES:
pH dependence:
Optimum pH is 4-6;
CC -I- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
subunits, the proteolytic products.
CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
as a proheparanase. In lysosomes, it is processed into the active
form, the heparanase. The uptake or internalisation of
proheparanase is mediated by HSPGs. Heparin appears to be a
competitor and retain proheparanase in the extracellular medium.
CC -I- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
weakly expressed in lymph node, thymus, peripheral blood
leukocytes, bone marrow, endothelial cells, fetal liver and tumor
tissues.
CC -I- PTM: Proteolytically processed. The cleavage of the 65 kDa form
leads to the generation of a linker peptide, 8 kDa and 50 kDa
product. The active form, the 8/50 kDa heterodimer, is resistant
to degradation. Complete removal of the linker peptide appears to
be a prerequisite to the complete activation of the enzyme.


```
CC -I- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.

Query Match      100.0%; Score 2020; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.3e-154;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 60
DB 158 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 217
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 218 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 277
QY 121 MKLSFLLKAGEVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQVQVVESTRPGK 180
DB 278 MKLSFLLKAGEVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQVQVVESTRPGK 337
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 240
DB 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 397
QY 241 FDPPLDYWLSLLPKLVGTVKVLMASSVQGSKRRKLRVYLHCTNTDNPYKEGDLTLVAINL 300
DB 398 FDPPLDYWLSLLPKLVGTVKVLMASSVQGSKRRKLRVYLHCTNTDNPYKEGDLTLVAINL 457
QY 301 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDQDTLPPLMEKPLR 360
DB 458 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDQDTLPPLMEKPLR 517
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 518 PGSSGLGPAFSYFFVIRNAKVAACI 543

RESULT 2
Q333X9_9RODE
ID Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      85.3%; Score 1724; DB 2; Length 574;
Best Local Similarity 84.5%; Pred. No. 2.9e-130;
Matches 325; Conservative 30; Mismatches 31; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 60
DB 189 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 248
```

```
Matches 326; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 60
DB 189 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 248
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 249 ISWELGNEPNSFWKAHISIDGLQGEDYIEHKLRLKSTLKNVKLYGPDVGQPRRTAK 308
QY 121 MKLSFLLKAGEVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQVQVVESTRPGK 180
DB 309 LLRSFLKAGEVIDSVTHHHYLLNGRIATKEDFLSPVDLDTFILSVQKILQVVESTRPGK 368
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 240
DB 369 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 428
QY 241 FDPPLDYWLSLLPKLVGTVKVLMASSVQGSKRRKLRVYLHCTNTDNPYKEGDLTLVAINL 300
DB 429 FEPLPDYWLSLLPKLVGSKVLMARVKGPDGRSKLRVYLHCTNINHPRYQEGDLTLVAINL 488
QY 301 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDQDTLPPLMEKPLR 360
DB 489 YNVTKHLKLPYQLFNKPVDKYLVKPLGPGGLLSKSVQLNGLTLKMVDQDTLPALTEKPLR 548
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 549 PGSSGLGPAFSYFFVIRNAKVAACI 574

RESULT 3
Q333X8_9RODE
ID Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48EBFE7CD0BCB34 CRC64;

Query Match      85.3%; Score 1723; DB 2; Length 574;
Best Local Similarity 84.2%; Pred. No. 3.5e-130;
Matches 325; Conservative 30; Mismatches 31; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 60
DB 189 KKFKNSTYSSRSVDVLYTFANCSDLIFGLNALLRTADLOWNSSNAQLLLDYCSSKGYN 248
```


[illegible]

RESULT 6

HPSE_BOVIN		STANDARD;	PRT; 545 AA.
ID	HPSE_BOVIN		
Q9MYX0;			
DT	11-OCT-2005,	integrated into UniProtKB/Swiss-Prot.	
DT	01-JUN-2001,	sequence version 2.	
DT	07-MAR-2006,	entry version 15.	
DE	Heparanase precursor (EC 3.2.-.-)	[Contains: Heparanase 8 kDa subunit;	
DE	Heparanase 50 kDa subunit].		
GN	Name=HPSE;		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;		
OC	Pecora; Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.		
RC	TISSUE=Placenta;		
RC	MEDLINE=21176669; PubMed=11277877;		
Kizaki K., Nakano H., Nakano H., Takahaashi T., Imai K., Haehizume K.;			
RT	"Expression of heparanase mRNA in bovine placenta during gestation."		
RL	Reproduction 121:573-580(2001).		
-!	FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans. Also implicated in the extravasation of leukocytes and tumor cell lines. Contributes to metastasis and angiogenesis (By similarity).		
-!	ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower extent, by heparin, sulfamin and EDTA. Activated by calcium and magnesium (By similarity).		
-!	SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa subunits, the proteolytic products (By similarity).		
-!	SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted. Secreted, internalised and transferred to late endosomes/lysosomes as a proheparanase. In lysosomes, it is processed into the active form, the heparanase. The uptake or internalisation of proheparanase is mediated by HSPGs. Heparin appears to be a competitor and retain proheparanase in the extracellular medium (By similarity).		
-!	TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the kidney, lung, spleen and uterus.		
-!	PTM: Proteolytically processed. The cleavage of the 65 kDa form leads to the generation of a linker peptide, 8 kDa and 50 kDa product. The active form, the 8/50 kDa heterodimer, is resistant to degradation. Complete removal of the linker peptide appears to be a prerequisite to the complete activation of the enzyme (By similarity).		
-!	PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears to be essential for its solubility (By similarity).		
-!	SIMILARITY: Belongs to the glycosyl hydrolase 79 family.		
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms Distributed under the Creative Commons Attribution-NoDerivs License		
EMBL; AF281160; AAF87301.2; -; mRNA.			
DR InterPro; IPR005199; Glyco_hydro_79_N.			
PFam; PF03662; Glyco_hydro_79n; 1.			
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane; Signal.			
KW SIGNAL	1	37	By similarity.
FT CHAIN	38	111	Heparanase 8 kDa subunit (By similarity). /FTId=PRO_000042256.
FT PROPEP	112	159	Linker peptide.

RP SUBUNITS.
RC STRAIN=FVB; TTSSU=Embryo;
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;
RT "Cloning, expression, and purification of mouse heparanase.";
RL Protein Expr. Purif. 26:425-431(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M30925200;
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,
RA Li J.-B.;
RT "Processing of macromolecular heparin by heparanase.";
RL J. Biol. Chem. 278:35152-35158(2003).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Aweller R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mortagi-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
extracellular matrix-degrading enzyme. Cleaves heparan sulfate
proteoglycans (HSPGs) into heparan sulfate side chains and core
proteoglycans. Also implicated in the extravasation of leukocytes
and tumor cell lines. Contributes to metastasis and angiogenesis
(By similarity).
CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and
magnesium (By similarity). Inhibited by laminarin sulfate and, to
a lower extent, by heparin and sulfamin.
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:
pH dependence:
Optimum pH is 5;
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
subunits, the proteolytic products.
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes

CC as a proheparanase. In lysosomes, it is processed into the active
form, the heparanase. The uptake or internalisation of
proheparanase is mediated by HSPGs. Heparin appears to be a
competitor and retain proheparanase in the extracellular medium
(By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
leads to the generation of a linker peptide, 8 kDa and 50 kDa
product. The active form, the 8/50 kDa heterodimer, is resistant
to degradation. Complete removal of the linker peptide appears to
be a prerequisite to the complete activation of the enzyme (By
similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
to be essential for its solubility.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF395507; AAQ15188.1; -; mRNA.
DR EMBL; AY077467; AAL76083.1; -; mRNA.
DR EMBL; AY151051; AAN41636.1; -; mRNA.
DR EMBL; AK040471; BAC30600.1; -; mRNA.
DR EMBL; AK154628; BAE32725.1; -; mRNA.
DR Ensembl; ENSMUSG00000035273; Mus musculus.
DR MGI; MGI:1343124; Hpsc.
DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
DR Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
Magnesium; Membrane; Signal.
FT SIGNAL 1 27 By similarity.
FT CHAIN 28 101 Heparanase 8 kDa subunit.
FT PROPEP 102 149 /FTid=PRO_0000042263.
FT CHAIN 150 535 /FTid=PRO_0000042264.
FT REGION 150 154 Heparanase 50 kDa subunit.
FT REGION 262 272 Heparin/HS-binding (By similarity).
FT ACT_SITE 217 217 Heparin/HS-binding (By similarity).
FT ACT_SITE 335 335 Proton donor (Potential).
FT CARBOHYD 154 154 Nucleophile (Potential).
FT CARBOHYD 192 192 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 209 209 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 230 230 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 451 451 N-linked (GlcNAc...) (Potential).
FT CONFLICT 206 206 K -> R (in Ref. 3).
FT CONFLICT 212 212 W -> S (in Ref. 3).
FT CONFLICT 230 232 NGS -> DGL (in Ref. 1, 2 and 4).
FT CONFLICT 335 335 E -> K (in Ref. 3).
FT CONFLICT 342 342 G -> A (in Ref. 3).
FT CONFLICT 455 455 Y -> H (in Ref. 1, 2 and 4).
FT CONFLICT 531 531 V -> I (in Ref. 1, 2 and 4).
SQ SEQUENCE 535 AA; 60050 MW; AF19E28B7CD03F7B CRC64;
Query Match 82.2%; Score 1661; DB 1; Length 535;
Best Local Similarity 80.8%; Pred. No. 3, 2e-125;
Matches 312; Conservative 33; Mismatches 41; Indels 0; Gaps 0;
QY 1 KKFKNSTYSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60
Db 150 KEFKNSTYSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 209
QY 61 ISWELGNEPNSFLKADIFINGSQLGDFQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db 210 ISWELGNEPNSFWKKAHILINGSQLGDFVELHKLQSFQNAKLYGPDIGQPRGKTVK 269
QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDFITSSVQKVFQVFWESTRPGK 180
Db 270 LKRSFLKAGGEVIDSVTHHYLYNGRTATREDFLSSVDLDTFLSVQKILKVTKEIFPK 329
QY 181 KWLIGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFFGAGNYHLVDEN 240

Db 330 KWLGETSSAYGGAPILLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDEN 389
QY 241 FDPDPYWLSSLFKKLVGTVMASVQSGKRKRLRVYLHCTNTDNPRYKGGDLTYAINL 300
Db 390 FEPLDPYWLSSLFKKLVGPRVLSRVKGPDRSKLVRVYLHCTNVVPRYQGGDLTYVNL 449
QY 301 HNTYKYLRLPYPPSPNQVDKYLRLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360
Db 450 HNTYKYLKVPPLFRFPVDYLLKPSGPDGLLSKSVQNLGQILKMWDEQTLPALTEKPLP 509
QY 361 PGSSGLGPAFSPSYFFVIRNAKVAACI 386
Db 510 AGSALSPLAFSYGFFVIRNAKVAACI 535
RESULT 8
HPSE RAT
ID HPSE RAT STANDARD; PRT; 536 AA.
AC Q71RPL; Q90ZF8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).
GN Name=Hps; Synonyms=Hep;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)_TaxID=10116;
RP NUCLEOTIDE SEQUENCE [MRNA].
RC TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis";
RL Nat. Med. 5:803-809(1999).
RN (2)
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yangishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line";
RL J. Biol. Chem. 277:32459-32465(2002).
CC !- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC !- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC !- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC !- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC !- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC !- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC !- SIMILARITY: Belongs to the glucosyl hydrolase 79 family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF59508; AAQ15189.1; -, mRNA.
DR EMBL; AF184967; AAF04563.1; -, mRNA.
DR RGD; 61969; Hps.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79a; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 Linker peptide (By similarity).
FT CHAIN 151 536 Heparanase 50 kDa subunit.
FT REGION 151 155 /FTid=PRO_000042268.
FT REGION 263 273 Heparin/HS-binding (By similarity).
FT ACT_SITE 218 218 Proton donor (Potential).
FT CARBOHYD 336 336 Nucleophile (Potential).
FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C43E04CF536EA4D CRC64;
Query Match 81.5%; Score 1646; DB 1; Length 536;
Best Local Similarity 80.1%; Pred. No. 5.3e-124;
Matches 309; Conservative 36; Mismatches 41; Indels 0; Gaps 0;
QY 1 KKPKNSTYSRSSVDLVTFYFACSGDLDFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60
Db 151 REFKNSTYSRSSVDMLYSPAKCSRLDLIFGLNALLRTPDLRWSSNAQLLDYCSSKGYN 210
QY 61 ISWELGNPNPSFLKKADIFINGSQGLGDFDTQLHLKLRKSTFNKAKLYGPDVGQPRRTAK 120
Db 211 ISWELGNPNPSFWKKAHISIDGLQLGDFVELHKLQKSAFQNAKLYGPDIGQPRGKTVK 270
QY 121 MLKSFLLKAGEVDSVTWHYLLNGRTATREDFLNPVLDLFISSVQKVFQVESTRGK 180
Db 271 LLRSFLKAGEVDSLTWHYLLNGRVATKEDFLSSVDLDFILSVQKILKVKEMTEPK 330
QY 181 KWLGETSSAYGGAPILLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDEN 240
Db 331 KWLGETSSAYGGAPILLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNYHLVDEN 390
QY 241 FDPDPYWLSSLFKKLVGTVMASVQSGKRKRLRVYLHCTNTDNPRYKGGDLTYAINL 300
Db 391 FEPLDPYWLSSLFKKLVGPRVLSRVKGPDRSKLVRVYLHCTNVVPRYQGGDLTYVNL 450
QY 301 HNTYKYLRLPYPPSPNQVDKYLRLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360
Db 451 HNTYKYLKVPPLFRFPVDYLLKPSGPDGLLSKSVQNLGQILKMWDEQTLPALTEKPLP 510
QY 361 PGSSGLGPAFSPSYFFVIRNAKVAACI 386
Db 511 AGSALSPLAFSYGFFVIRNAKVAACI 536
RESULT 9
Q33X5_SPAJD
ID Q33X5_SPAJD PRELIMINARY; PRT; 558 AA.
AC Q33X5;
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).


```
Db 370 TYTPGKKIWLEGVVTTTSGAGTNNLSDSYAAGFLWNTLGLMLANQGDIVVIRHSFFDHGYN 429
QY 235 HLVDENFDPLDPDYWLSLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 285
Db 430 HLVDQNFNPLDPDYWLSLLYKRLGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHN 489
QY 286 PRYKEGDLTYAINLHNVTYKYLRLPYFPFSNKQVDKYLLRPLGPHGLSKSVOLNGLTKM 345
Db 490 HNTVRSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQSGLSKSVQNLGQPLVM 549
QY 346 VDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKVAAC 385
Db 550 VDDGTLPELKPRLRAGRTLIVPVTMGFFVVKVNNALAC 589

RESULT 13
Q2M1H9 HUMAN
ID Q2M1H9_HUMAN PRELIMINARY; PRT; 592 AA.
AC Q2M1H9
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 1.
DE Heparantase 2.
GN Name=HPSE2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones; PubMed12477932; DOI=10.1073/pnas.242603899;
RX MEDLINE=22388257; PubMed12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shen C.M., Schuler G.D.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan A., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Tohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu L.J., Hulyk S.W.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RG NIH MGC Project;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC112356; AAI12357.1; -; mRNA.
CC SEQUENCE 592 AA; 66610 MW; 946899E1C2A74359F CRC64;

Query Match 48.4%; Score 978; DB 2; Length 592;
Best Local Similarity 48.8%; Pred. No. 5,7e-70;
Matches 195; Conservative 60; Mismatches 129; Indels 16; Gaps 4;

QY 1 KKFKNSTYSR-----SSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQILLDYCS 55
Db 191 EQFSN-TYSNLIILARSGLKYNFADCSGLHLIFALNALRNPNNWNSSSALLSKYSA 249
```

```
QY 56 SKGYNISWELGNEPNSFLKKADIFINGSQLEDFTQLHKLRLK-STFKNAKLYGPDVGOP 114
Db 250 SKKYNISWELGNEPNNYTMHGRAVNGSQLGKDYIQLSKLLQPIRIYSRSLYGPNGIRP 309
QY 115 RRKTAKMLKSLKAGGEVIDSVTHHHYHNGRTATREDPLNDPDLIDIFISSVQKVFQVVE 174
Db 310 RKNVIALLDGPMKVGASTVDATWQHCHYIDGRVVKVMDFLKTRLLDLSQIRKIQKVN 369
QY 175 STRFGKVKWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNY 234
Db 370 TYTPGKKIWLEGVVTTTSGAGTNNLSDSYAAGFLWNTLGLMLANQGDIVVIRHSFFDHGYN 429
QY 235 HLVDENFDPLDPDYWLSLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 285
Db 430 HLVDQNFNPLDPDYWLSLLYKRLGPKVLAVHVGAGLQRPGRVIRDKLRIYAHCTNNHN 489
QY 286 PRYKEGDLTYAINLHNVTYKYLRLPYFPFSNKQVDKYLLRPLGPHGLSKSVOLNGLTKM 345
Db 490 HNTVRSITLFIINLHRSRKKIKLAGTLRDKLVHQYLLQPYQSGLSKSVQNLGQPLVM 549
QY 346 VDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKVAAC 385
Db 550 VDDGTLPELKPRLRAGRTLIVPVTMGFFVVKVNNALAC 589

RESULT 14
Q4TB80 TETNG
ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAP7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG0003869001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Basilya C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J., Robinson-Rechavi M.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype."
RL Nature 431:946-957(2004).
[2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CAAB01007180; CAF89852.1; -; Genomic DNA.
CC SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
FT NON TER 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
```

Query Match 43.3%; Score 874; DB 2; Length 597;
Best Local Similarity 45.2%; Pred. No. 1.4e-61;
Matches 180; Conservative 61; Mismatches 141; Indels 16; Gaps 5;
CC 3 FKNSTYSRSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLDYCSKSGYNISWELGNE 62
DB 200 FSHSRIAR-SLDKLYNFADACAGLHLILGLNALHRNPDHSWNTSTLSLLKYSAGKKNIS 258
QY 63 WELGNEPNSGLKADIFINGSQGLDFIQLHKLRLK-STFKNAKLYGPDVQPRRTAKM 121
DB 259 WELGNEPNAYSRMVGHAVNSQAQDYTKURLTQSVRYRAQLYGPNAGRPNKNAALL 318
QY 122 LKSFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTREPGKK 181
DB 319 LDEFMKVTGVVDVATWQHYMDGRKKVEDEFLKTRLLDITLTLQLSKVTKVNVHTTEGKK 378
QY 182 VWLGETSSAYGGAPLLSDTFAAGFWMLDKLGUSARMGIEVVMR-----QVFFAGNHYLV 237
DB 379 VWLGGIGPAWTGMSNLSDTFAAGFLVNTLGMAMQGIDVLLRQAVQVHTNKQSVLF 438
QY 238 DENEOP-LPDYWLFLFKLVGKTVLMASVQSKRR-----KLRVYLHCTNTDNP 287
DB 439 LQMFVPSFPDYFSLVFKRLGVKVLAVRAGLQRPQGRVIRDKLRIYARCTSYSNHN 498
QY 288 YKEGDLTLYAINLHNVTYLRPLPYPSNKKQVDKYLLRPLGPHGLLSKSVOLNGLTLKQVD 347
DB 499 YVRGSIITIIILNRSRKKIKLAGLIRNNIVHQLQPYGADGLRAKHVQNLGEKLLMAD 558
QY 348 DQTLPLMEKPLRPGSSGLPAPSFYSFFVIRNAKVAAC 385
DB 559 NETFPELKPRTLRAGRTIAMPPTGTFYVYIKINAYAC 596

RESULT 15
Q4TCG8 TETNG PRELIMINARY; PRT; 255 AA.
AC Q4TCG8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0001168001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallion O., Aury J.-M., Brunet P., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segreus B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionmont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an

EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; CAAE01003783; CAF88054.1; -; Genomic_DNA.
DR NON TER 1 1
FT NON TER 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;
Query Match 36.8%; Score 742.5; DB 2; Length 255;
Best Local Similarity 56.7%; Pred. No. 1.9e-51;
Matches 144; Conservative 37; Mismatches 56; Indels 17; Gaps 3;
QY 9 SRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLDYCSKSGYNISWELGNE 68
DB 1 SETTVDDQLHAFANCSGLDLVFGLNALLRTADNRWNSNARSLLRYCEARRYHMSWELGNE 60
QY 69 PNSFLKKADIFINGSQGLDFIQLHKLRLKSTF-KNAKLYGPDVQPRRTAKMLKSLK 127
DB 61 PNSYEKKAGLRDGRQLGEDFTVLRKILRESRFYRDAGLFGPDVQPRDHRIDILSGFLQ 120
QY 128 AGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTREPGKKVWLGET 187
DB 121 SGAEAVDACTWHHYLYDGREASLEDFLDPVLDTLREKIGEVLEEVHQSVPKPVWLGET 180
QY 188 SSAYGGG-----APLLSDTFAAG-FMWLDKLGLSARMGIEVVMRQVFFGA 231
DB 181 SSATGAEPGRCTHSSQDSCEFAFRSDQAPLGTFRWLDKLGLAATLGLLVMRQVLGA 240
QY 232 GNYHLVDENFDPLP 245
DB 241 GSTHLMDDNLDPLP 254

Search completed: June 5, 2006, 12:20:13
Job time : 96.7733 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 76.9552 Seconds
(without alignments)
2293.354 Million cell updates/sec

Title: US-10-645-659A-1
Perfect score: 2020
Sequence: 1 KFKNSTYRSRSDVLYTFA.....LPAFSYFFVIRNAKVAACI 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*
10: Geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2020	100.0	386	8	Adr88207 Human mat
2	2020	100.0	386	8	Adt78174 45kDa sub
3	2020	100.0	386	9	Ady27057 Heparanas
4	2020	100.0	386	9	Adz18995 Human hep
5	2020	100.0	386	9	Aea42423 Human mat
6	2020	100.0	460	9	Agv27061 Heparanas
7	2020	100.0	432	9	Adz18996 Hep106 co
8	2020	100.0	495	9	Adz18999 Hep109 co
9	2020	100.0	501	9	Adz19000 HepG3 co
10	2020	100.0	507	9	Adz19005 HepG6 co
11	2020	100.0	508	9	Ady27058 Human ina
12	2020	100.0	526	9	Adz19006 HepHyalur
13	2020	100.0	527	5	Abb07815 Chicken s
14	2020	100.0	527	7	Adw02018 Chimeric
15	2020	100.0	527	9	Adz19004 HepG4 co
16	2020	100.0	543	2	Aay17082 Human hep
17	2020	100.0	543	4	Aab86206 Human hep
18	2020	100.0	543	7	Adz18950 Human dis
19	2020	100.0	543	8	Adk52086 Human ato
20	2020	100.0	543	8	Adm48759 Human hpa
21	2020	100.0	543	8	Adn05074 Antipsori
22	2020	100.0	543	8	Adn04902 Antipsori
23	2020	100.0	543	8	Adq80372 Heparanas

24	2020	100.0	543	8	ADR88210	Human pre
25	2020	100.0	543	8	ADP25079	PRO polyp
26	2020	100.0	543	8	ADT78177	Human hep
27	2020	100.0	543	9	ADY27036	Human hep
28	2020	100.0	543	9	AEA42426	Human hep
29	2020	100.0	545	6	ABP56822	Human hep
30	2020	100.0	545	7	ADE16012	G-coupled
31	2020	100.0	545	8	ADL93951	Human G-c
32	2020	100.0	556	9	ADZ19010	Heparanas
33	2020	100.0	570	9	ADZ19008	Heparanas
34	2020	100.0	588	2	AAZ30124	A human p
35	2017	99.9	530	2	AAZ34173	Human pre
36	2017	99.9	543	2	AAV02345	A human h
37	2017	99.9	543	3	AAZ57590	Human hep
38	2017	99.9	543	3	AAZ08849	Amino aci
39	2017	99.9	543	3	AAZ52990	Human hep
40	2017	99.9	543	4	AAZ97635	Human hep
41	2017	99.9	543	5	ABB07813	Human hep
42	2017	99.9	543	7	ADG88800	Human hpa
43	2017	99.9	543	8	ADL16379	Human hep
44	2017	99.9	543	8	ADM48716	Human hpa
45	2017	99.9	543	9	AEA42466	Human hep

ALIGNMENTS

RESULT 1
ADR88207
ID ADR88207 standard; protein; 386 AA.
XX
AC ADR88207;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human mature heparanase 45 kDa major subunit.
XX
KW Targeted drug delivery ; inflammatory disorder; wound; scar;
KW vasculopathy; autoimmune disorder; cancer; angiogenesis;
KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;
KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;
KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;
KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.
XX
OS Homo sapiens.
XX
PN US2004170631-A1.
XX
PD 02-SEP-2004.
XX
PF 28-NOV-2003; 2003US-00722502.
XX
PR 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
PR 22-AUG-2003; 2003US-00645659.
XX
PA (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (VDEL/) VAN GELDER J M.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
PI Friedmann Y;
XX

DR WPI; 2004-625084/60.
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
PT useful for treating heparanase-associated conditions such as inflammation
PT or cancer, comprises administering a drug and an anti-heparanase antibody
PT complex.
XX
PS Claim 2; SEQ ID NO 1; 58pp; English.
XX
CC The invention relates to a method of targeted drug delivery to a tissue
CC of a patient, the tissue expressing heparanase. The method comprises
CC providing a complex of a drug directly or indirectly linked to an anti-
CC heparanase antibody, and administering the complex to the patient. In the
CC targeted drug delivery, the antibody comprises an epitope of a heparanase
CC capable of specifically binding to at least one epitope of a heparanase
CC protein. The composition and methods of the invention are useful for
CC diagnosing, preventing or treating conditions associated with heparanase
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell
CC proliferation, invasion of circulating tumour cells and metastatic
CC disease), for purifying heparanase, or for developing drugs for those
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and
CC multiple myeloma. The solid cancer is selected from tumours in lip and
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's
CC sarcoma. The present sequence is the 45 kDa major subunit of human mature
CC heparanase.
XX
SQ Sequence 386 AA;

Query Match 100.0%; Score 2020; DB 8; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
DB 1 KKFKNSTYSSRSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
QY 61 ISWELGNPNFLKKADIFINGSQLGDFIQLHLKLRKFTFNKALYGPDPVQPRRTAK 120
DB 61 ISWELGNPNFLKKADIFINGSQLGDFIQLHLKLRKFTFNKALYGPDPVQPRRTAK 120
QY 121 MLKSLFKAGGEVIDSVTHHYLYNGRTATREDPLNDVLDIFISSQVKVQVVESTRPGK 180
DB 121 MLKSLFKAGGEVIDSVTHHYLYNGRTATREDPLNDVLDIFISSQVKVQVVESTRPGK 180
QY 181 KYWLGETTSAYCGGAPLLSDTTPAAGFMWLDKGLSARMGLEVMVRQVFCAGNYHLVDEN 240
DB 181 KYWLGETTSAYCGGAPLLSDTTPAAGFMWLDKGLSARMGLEVMVRQVFCAGNYHLVDEN 240
QY 241 FDPDLYWLSLLFKKLVGTKVLMASVQSGKRRKRLRYLHCTNTDNPYKRGDITLYAINL 300
DB 241 FDPDLYWLSLLFKKLVGTKVLMASVQSGKRRKRLRYLHCTNTDNPYKRGDITLYAINL 300
QY 301 HNVTKYRLPYFPFNQVDKYLRLPGHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360
DB 301 HNVTKYRLPYFPFNQVDKYLRLPGHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360

QY 361 PGSSGLPAFSYSFFVIRNAKVAACI 386
DB 361 PGSSGLPAFSYSFFVIRNAKVAACI 386
RESULT 2
ADT78174
ID ADT78174 standard; protein; 386 AA.
XX AC
XX ADT78174;
XX AC
XX 13-JAN-2005 (first entry)
XX
DE 45kDa subunit of mature processed human heparanase dimer.
XX
KW Antibody; epitope; heparanase; pathological condition; angiogenesis;
KW cell proliferation; cancerous condition; tumour cell invasion;
KW metastatic disease; heparanase-related disorder; inflammatory disorder;
KW wound; scar; vasculopathy; autoimmune condition; renal disease;
KW cystostatic; antiinflammatory; vulnery; antiarteriosclerotic;
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.
OS Homo sapiens.
XX
PN US2004213789-A1.
XX
PD 28-OCT-2004.
XX
PF 22-AUG-2003; 2003US-00645659.
XX
PR 02-SEP-1997; 97US-00922170.
PR 01-MAY-1998; 98US-00071739.
PR 04-NOV-1998; 98US-00186200.
PR 19-FEB-2003; 2003US-00368044.
XX
XX (YACO/) YACOBY-ZEEVI O.
PA (PERE/) PERETZ T.
PA (MIRO/) MIRON D.
PA (SHLO/) SHLOMI Y.
PA (PECK/) PECKER I.
PA (AYAL/) AYAL-HERSHKOVITZ M.
PA (FEIN/) FEINSTEIN E.
PA (GELD/) GELDER J M V.
PA (VLOD/) VLODAVSKY I.
PA (FRIE/) FRIEDMANN Y.
XX
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;
PI Friedmann Y;
XX
XX WPI; 2004-774790/76.
XX
PT New neutralizing monoclonal anti-heparanase antibodies, useful for
PT detecting, treating or preventing cancer, inflammatory or autoimmune
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.
XX
PS Claim 5; SEQ ID NO 1; 68pp; English.
XX
CC The invention relates to an isolated antibody or antibody portion capable
CC of specifically binding to or elicited by at least one epitope of a
CC heparanase protein, where the heparanase protein is at least 60%
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and
CC where at least one epitope comprises a sequence at least 70% homologous
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)
CC a hybridoma cell line comprising a cell line for producing the monoclonal
CC antibody, (b) a method for detecting, treating or preventing a
CC pathological condition or a heparanase-related disorder or condition in a
CC subject, (c) a method for monitoring the state of a heparanase-related
CC disorder or condition in a subject, and (d) a pharmaceutical composition
CC comprising the isolated anti-heparanase antibody or antibody portion and
CC a pharmaceutical carrier. The antibody, methods, and composition are
CC useful for detecting, treating, preventing or monitoring a pathological
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,
 CC or prostate cancer), minor cell proliferation, invasion of circulating
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of
 CC mature processed human heparanase dimer.

XX
 SQ Sequence 386 AA;

Query Match 100.0%; Score 2020; DB 8; Length 386;
 Best Local Similarity 100.0%; Pred. No. 4.9e-202;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
 DB 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
 QY 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLLRKSTFKNAKLYGPDVGQPRRTAK 120
 DB 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLLRKSTFKNAKLYGPDVGQPRRTAK 120
 QY 121 MLKSFLLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVESTPCK 180
 DB 121 MLKSFLLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVESTPCK 180
 QY 181 KWLGTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
 DB 181 KWLGTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
 QY 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSKRRKRLRVYLHCTNTDNPYKSGDLTYAINL 300
 DB 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSKRRKRLRVYLHCTNTDNPYKSGDLTYAINL 300
 QY 301 HNVTKYLRPLPYPSNKKQVDKYLLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360
 DB 301 HNVTKYLRPLPYPSNKKQVDKYLLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360
 QY 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386
 DB 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386

RESULT 3

ADY27057

ID ADY27057 standard; protein; 386 AA.

AC ADY27057;

XX ADY27057;

DT 05-MAY-2005 (first entry)

XX Heparanase inhibitor protein #1.

DE Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;

XX neurological disease; viral infection; infection; cytostatic;

KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;

KW heparanase modulator; enzyme purification.

XX Homo sapiens.

OS WO2005016227-A2.

XX 24-FEB-2005.

XX 12-AUG-2004; 2004WO-IL000744.

XX 14-AUG-2003; 2003US-0494800P.

PR 12-JAN-2004; 2004US-0535492P.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA

XX

PI Van-Gelder JM, Miron D;

XX WPI; 2005-182203/19.

DR

XX

PT Regulating heparanase activity, useful for treating heparanase-associated

PT diseases (e.g. cancer, inflammation, cardiovascular diseases,

PT neurological diseases or viral diseases) comprises modulating heparanase

PT activation.

XX

PS Claim 55; SEQ ID NO 33; 211pp; English.

XX

CC The invention relates to a method of regulating heparanase activity in a

CC tissue or regulating a biological process depending at least in part on

CC heparanase activity comprising modulating heparanase activation. The

CC invention also relates to methods of treating a heparanase- or heparin

CC binding protein-associated disease or disorder in a subject, a

CC pharmaceutical composition for use in the treatment of a heparanase-

CC associated disease or disorder comprising a therapeutic amount of an

CC agent capable of modulating heparanase activation and a pharmaceutical

CC carrier or diluent, a method of identifying a protease activator of

CC heparanase, a protease substrate mimetic comprising a peptide

CC representing a subset or all substrate residues or cleavage sites of

CC human heparanase or an equivalent non-human heparanase, a method of

CC producing active heparanase and a method of modulating an adhesion

CC activity of heparanase. The composition and methods are useful for

CC modulating heparanase activation and for treating heparanase-associated

CC diseases or disorders such as cancer, inflammation, cardiovascular

CC diseases, neurological diseases or viral infections. This sequence

CC represents a heparanase inhibitor protein used in the scope of the

CC invention.

XX

SQ Sequence 386 AA;

Query Match 100.0%; Score 2020; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 4.9e-202;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60

DB 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60

QY 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLLRKSTFKNAKLYGPDVGQPRRTAK 120

DB 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLLRKSTFKNAKLYGPDVGQPRRTAK 120

QY 121 MLKSFLLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVESTPCK 180

DB 121 MLKSFLLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVESTPCK 180

QY 181 KWLGTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

DB 181 KWLGTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

QY 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSKRRKRLRVYLHCTNTDNPYKSGDLTYAINL 300

DB 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSKRRKRLRVYLHCTNTDNPYKSGDLTYAINL 300

QY 301 HNVTKYLRPLPYPSNKKQVDKYLLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360

DB 301 HNVTKYLRPLPYPSNKKQVDKYLLRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360

QY 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386

DB 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386

RESULT 4

ADZ18995

ID ADZ18995 standard; protein; 386 AA.

XX ADZ18995;

AC

XX 16-JUN-2005 (first entry)

DT

XX Human heparanase consensus cleavage site #2.
DE
XX Enzyme engineering; heparanase; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX Homo sapiens.
OS
XX WO2005030962-A1.
PN
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
PF
XX 26-SEP-2003; 2003US-0506479P.
PR
XX 20-JAN-2004; 2004US-0537729P.
PR
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
PA
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
PI
XX WPI; 2005-273382/28.
DR
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Disclosure; SEQ ID NO 16; 65pp; English.
PS
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a human heparanase
CC consensus cleavage site used in the scope of the invention.
XX
XX Sequence 386 AA;
SQ
Query Match 100.0%; Score 2020; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSSRSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60
DB 1 KKFKNSTYSSRSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60
QY 61 ISWELGNEPNSFLKKADIFINGSOLGEDFTQLHKLKSTFKNAKLYGPDVGOPRRKTAK 120
DB 61 ISWELGNEPNSFLKKADIFINGSOLGEDFTQLHKLKSTFKNAKLYGPDVGOPRRKTAK 120
QY 121 MLKSFLLKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPKG 180
DB 121 MLKSFLLKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPKG 180
QY 181 KVMLGTSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 181 KVMLGTSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
QY 241 FDPDPDYWLSLFPKLVGCTKVLMAVQGSRRKRLVYLHCTNTDNPYKEGDTLYAINL 300
DB 241 FDPDPDYWLSLFPKLVGCTKVLMAVQGSRRKRLVYLHCTNTDNPYKEGDTLYAINL 300
QY 301 HNVTKYLRLPYPSNKQVDKYLRLPLGPHGLSKSVQLNGLTLLKWDQDQTLPLMEKPLR 360
DB 301 HNVTKYLRLPYPSNKQVDKYLRLPLGPHGLSKSVQLNGLTLLKWDQDQTLPLMEKPLR 360
QY 361 PGSSGLGLPAFSYFFVIRNAKVAACI 386
DB 361 PGSSGLGLPAFSYFFVIRNAKVAACI 386

RESULT 5
AEA42423
ID AEA42423 standard; protein; 386 AA.
XX
AC AEA42423;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.
XX
KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
KW angiogenesis disorder; cancer; tumor; metastasis.
XX
OS Homo sapiens.
XX
PN AU2004201462-A1.
XX
PD 06-MAY-2004.
XX
PF 08-APR-2004; 2004AU-00201462.
XX
PR 08-APR-2004; 2004AU-00201462.
XX
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
PI Feinstein E;
XX
DR WPI; 2005-173343/19.
XX
PT Novel isolated antibody capable of specifically binding to epitope of
PT heparanase protein, useful for preventing and treating heparanase-related
PT disorder such as inflammatory disorder, scars, autoimmune conditions or
PT angiogenesis.
XX
PS Claim 2; SEQ ID NO 1; 260pp; English.
XX
CC The invention relates to an isolated antibody or its portion (I) capable
CC of specifically binding to an epitope of a heparanase protein. Also
CC described: (1) a cell line (II) for producing a monoclonal antibody or
CC its portion, comprising a cell line for producing (I); (2) a
CC pharmaceutical composition comprising (I) and a carrier; and (3) an
CC affinity medium (III) for binding human heparanase polypeptides,
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)
CC useful for treating a subject suffering from a pathological condition,
CC which involves administering (I) to the subject. (I) is useful for
CC preventing and treating heparanase-related disorder or condition chosen
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor
CC cell proliferation, invasion of circulating tumor cells and metastatic
CC disease. (I) is useful for detecting the presence of heparanase
CC polypeptide in a sample. (I) is useful for detecting heparanase-related
CC disease or condition in a subject such as vertebrate, preferably mammal
CC e.g., human. The heparanase-related disorder or condition further
CC includes renal disease or disorder chosen from diabetic nephropathy,
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome
CC and renal cell carcinoma. The present sequence represents the 45 kDa
CC subunit of the human mature processed heparanase dimer, which is used in
CC the exemplification of the present invention.
XX
SQ Sequence 386 AA;

Query Match 100.0%; Score 2020; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 4.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60

Db 1 KKFKNSTYSRSSVDVLYTFANCSDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60
QY 61 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db 61 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
QY 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKVQVVESTPRGK 180
Db 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKVQVVESTPRGK 180
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
Db 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
QY 241 FDPDPYWLSSLFKKLVGTGKVMASVQGSKRRKRLRVYLHCTNTDNPRYKEGDLTYAINL 300
Db 241 FDPDPYWLSSLFKKLVGTGKVMASVQGSKRRKRLRVYLHCTNTDNPRYKEGDLTYAINL 300
QY 301 HNVTKYLRLPYPFNSKNQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR 360
Db 301 HNVTKYLRLPYPFNSKNQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR 360
QY 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386
Db 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386

RESULT 6
ADY27061
ID ADY27061 standard; protein; 460 AA.
XX ADY27061;
XX AC
XX DT
XX DT
XX DE Heparanase inhibitor protein #4.
XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.
XX OS Homo sapiens.
XX PN WO2005016227-A2.
XX PD 24-FEB-2005.
XX PF 12-AUG-2004; 2004WO-11000744.
XX PR 14-AUG-2003; 2003US-0494800P.
XX PR 12-JAN-2004; 2004US-0535492P.
XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PI Van-Gelder JM, Miron D;
XX WPI; 2005-182203/19.
XX DR
XX PT Regulating heparanase activity, useful for treating heparanase-associated
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT neurological diseases or viral diseases) comprises modulating heparanase
PT activation.
XX PS Disclosure; SEQ ID NO 37; 21pp; English.
XX CC

CC associated disease or disorder comprising a therapeutic amount of an
CC agent capable of modulating heparanase activation and a pharmaceutical
CC carrier or diluent, a method of identifying a protease activator of
CC heparanase, a protease substrate mimetic comprising a peptide
CC representing a subset or all substrate residues or cleavage sites of
CC human heparanase or an equivalent non-human heparanase, a method of
CC producing active heparanase and a method of modulating an adhesion
CC activity of heparanase. The composition and methods are useful for
CC modulating heparanase activation and for treating heparanase-associated
CC diseases or disorders such as cancer, inflammation, cardiovascular
CC diseases, neurological diseases or viral infections. This sequence
CC represents a heparanase inhibitor protein used in the scope of the
CC invention.
XX
SQ Sequence 460 AA;
Query Match 100.0%; Score 2020; DB 9; Length 460;
Best Local Similarity 100.0%; Pred. No. 6.4e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSRSSVDVLYTFANCSDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60
Db 75 KKFKNSTYSRSSVDVLYTFANCSDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 134
QY 61 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db 135 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 194
QY 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKVQVVESTPRGK 180
Db 195 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKVQVVESTPRGK 254
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
Db 255 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 314
QY 241 FDPDPYWLSSLFKKLVGTGKVMASVQGSKRRKRLRVYLHCTNTDNPRYKEGDLTYAINL 300
Db 315 FDPDPYWLSSLFKKLVGTGKVMASVQGSKRRKRLRVYLHCTNTDNPRYKEGDLTYAINL 374
QY 301 HNVTKYLRLPYPFNSKNQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR 360
Db 375 HNVTKYLRLPYPFNSKNQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR 434
QY 361 PGSSGLGPAFSYSFFVIRNAKVAACI 386
Db 435 PGSSGLGPAFSYSFFVIRNAKVAACI 460
RESULT 7
ADZ18996
ID ADZ18996 standard; protein; 492 AA.
XX AC ADZ18996;
XX DT 16-JUN-2005 (first entry)
XX DE Hep106 construct protein.
XX KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX OS Synthetic.
XX PN WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.

```

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX PD WPI: 2005-273382/28.
XX DR N-PSDB; ADZ18997.
XX DR
XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX PT nucleotide sequence that encodes mammalian heparanase protein and has two
XX PT consensus cleavage sites located between specific nucleotide encoding
XX PT residues.
XX PS Example 2; SEQ ID NO 17; 65pp; English.
XX SQ
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepi06 construct protein
CC used in the scope of the invention.
XX SQ Sequence 492 AA;
Query Match 100.0%; Score 2020; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 7.2e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
Db 107 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 166
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db 167 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 226
QY 121 MLKSFLLKAGEVIDSVTWHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 180
Db 227 MLKSFLLKAGEVIDSVTWHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 286
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDEN 240
Db 287 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDEN 346
QY 241 FDPDPYWLSSLFVKLVGTGKVLMAVQGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300
Db 347 FDPDPYWLSSLFVKLVGTGKVLMAVQGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 406
QY 301 HNVTKYLRLPYFPNSKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360
Db 407 HNVTKYLRLPYFPNSKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 466
QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386
Db 467 PGSSGLGLPAFSYSFFVIRNAKVAACI 492
RESULT 8
ID ADZ18999
XX ADZ18999 standard; protein; 495 AA.
XX AC ADZ18999;
XX DT 16-JUN-2005 (first entry)
XX DE Hep109 construct protein.
XX KW Enzyme engineering; heparanase; hepi09; metastasis; autoimmune disease;
XX KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX KW immunosuppressive; enzyme.
XX

```

```

OS Synthetic.
XX WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
XX PR 20-JAN-2004; 2004US-0537729P.
XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX DR WPI: 2005-273382/28.
XX DR N-PSDB; ADZ18998.
XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX PT nucleotide sequence that encodes mammalian heparanase protein and has two
XX PT consensus cleavage sites located between specific nucleotide encoding
XX PT residues.
XX PS Example 2; SEQ ID NO 20; 65pp; English.
XX SQ
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepi09 construct protein
CC used in the scope of the invention.
XX SQ Sequence 495 AA;
Query Match 100.0%; Score 2020; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 7.2e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
Db 110 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 169
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db 170 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 229
QY 121 MLKSFLLKAGEVIDSVTWHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 180
Db 230 MLKSFLLKAGEVIDSVTWHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 289
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDEN 240
Db 290 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDEN 349
QY 241 FDPDPYWLSSLFVKLVGTGKVLMAVQGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300
Db 350 FDPDPYWLSSLFVKLVGTGKVLMAVQGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 409
QY 301 HNVTKYLRLPYFPNSKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360
Db 410 HNVTKYLRLPYFPNSKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 469
QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386
Db 470 PGSSGLGLPAFSYSFFVIRNAKVAACI 495
RESULT 9
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX

```


AC ADZ19000;
XX
DT 16-JUN-2005 (first entry)
XX
DE HepG3 construct protein.
XX
KW Enzyme engineering; heparanase; hepg3; metastasis; autoimmune disease;
inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
DR WPI; 2005-273382/28.
DR N-PSDB; ADZ19001.
XX
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
PS Example 2; SEQ ID NO 21; 65pp; English.
XX
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepg3 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 501 AA;

Query Match 100.0%; Score 2020; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 7.4e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
DB 116 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 175

QY 61 ISWELGNEPNSFLKKADIFINGSQGLDFIQLHKLRLKSTFKNAKLYGPDVQPPRRKTAK 120
DB 176 ISWELGNEPNSFLKKADIFINGSQGLDFIQLHKLRLKSTFKNAKLYGPDVQPPRRKTAK 235

QY 121 MLKSFLLKAGGEVIDSVTHHYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 180
DB 236 MLKSFLLKAGGEVIDSVTHHYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 295

QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 296 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 355

QY 241 FDPPLPWLSLLPKLVGTGTVLMASVQSKRRKRLRVYLHCTNTDNPYKGGDLTLYAINL 300
DB 356 FDPPLPWLSLLPKLVGTGTVLMASVQSKRRKRLRVYLHCTNTDNPYKGGDLTLYAINL 415

QY 301 HNYTKYLRPLYPFSNQVDKYLLRPLGPHGLLSKSVQNLGTLTKMWDDQTLPLMEKPLR 360
DB 416 HNYTKYLRPLYPFSNQVDKYLLRPLGPHGLLSKSVQNLGTLTKMWDDQTLPLMEKPLR 475

QY 361 PGSSGLGLPAFSYSPFVIRNAKVAACI 386
DB 476 PGSSGLGLPAFSYSPFVIRNAKVAACI 501

RESULT 10
ADZ19005
ID ADZ19005 standard; protein; 507 AA.
XX
AC ADZ19005;
XX
DT 16-JUN-2005 (first entry)
XX
DE HepG6 construct protein.
XX
KW Enzyme engineering; heparanase; hepg6; metastasis; autoimmune disease;
inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
DR WPI; 2005-273382/28.
DR N-PSDB; ADZ19003.
XX
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
PS Example 2; SEQ ID NO 26; 65pp; English.
XX
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepg6 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 507 AA;

Query Match 100.0%; Score 2020; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 7.5e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
DB 122 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 181

QY 61 ISWELGNEPNSFLKKADIFINGSQGLDFIQLHKLRLKSTFKNAKLYGPDVQPPRRKTAK 120
DB 182 ISWELGNEPNSFLKKADIFINGSQGLDFIQLHKLRLKSTFKNAKLYGPDVQPPRRKTAK 241

QY 121 MLKSFLLKAGGEVIDSVTHHYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 180
DB 242 MLKSFLLKAGGEVIDSVTHHYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 301

QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

Db	302	KVNLGETSSAYCGGAPLLSDTFAAGFMWLDKGLSARMGIEYVMRQVFFGAGNYHLVDEN	361
Qy	241	FDPLPDYWSLLFKKLVTGKVLMAVQSGSKRRKLRVYLHCTNTDNPRYKEGDLTLIYAINL	300
Db	362	FDPLPDYWSLLFKKLVTGKVLMAVQSGSKRRKLRVYLHCTNTDNPRYKEGDLTLIYAINL	421
Qy	301	HNVTKYLRLPYFPSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKWDDQDTLPPLMEKPLR	360
Db	422	HNVTKYLRLPYFPSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKWDDQDTLPPLMEKPLR	481
Qy	361	PGSSIGLPAFSYSFPIRNKVAACI	386
Db	482	PGSSIGLPAFSYSFPIRNKVAACI	507

RESULT 11	
ADY27058	
ID	ADY27058 standard; protein; 508 AA.
XX	
XX	ADY27058;
AC	
XX	
XX	05-MAY-2005 (first entry)
XX	
DE	Human inactive heparanase protein.
XX	
KW	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW	neurological disease; viral infection; infection; cytostatic;
KW	antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW	protease; enzyme; enzyme purification.
XX	
OS	Homo sapiens.
XX	
PN	WO2005016227-A2.
XX	
XX	24-FEB-2005.
PD	
XX	
XX	12-AUG-2004; 2004WO-IL000744.
XX	
PR	14-AUG-2003; 2003US-0494800P.
PR	12-JAN-2004; 2004US-0535492P.
XX	
XX	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA	
XX	
PI	Van-Gelder JM, Miron D;
XX	
XX	WPI; 2005-182203/19.
DR	
XX	
PT	Regulating heparanase activity, useful for treating heparanase-associated
PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT	neurological diseases or viral diseases) comprises modulating heparanase
PT	activation.
XX	
PS	Claim 257; SEQ ID NO 34; 211pb; English.

The invention relates to a method of regulating heparanase activity in a tissue or regulating a biological process depending at least in part on heparanase activity comprising modulating heparanase activation. The heparanase activity comprises modulating heparanase activation. The invention also relates to methods of treating a heparanase- or heparin binding protein-associated disease or disorder in a subject, a pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence represents a human inactive heparanase protein used in the scope of the

CC	invention.
XX	
SQ	Sequence 508 AA;
	Query Match 100.0%; Score 2020; DB 9; Length 508;
	Best Local Similarity 100.0%; Pred. No. 7.5e-202;
	Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 KKFKNSTYSRSSVDVLTFFANCSGLDLI FGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
Dd	123 KKFKNSTYSRSSVDVLTFFANCSGLDLI FGLNALLRTADLQWNSNAQLLLDYCSSKGYN 182
Qy	61 ISMELGNPNLSFLKKADIFINGSOLGEDFIQLHKLLRKSTFKNAKLXGPDVGQPRRKTAK 120
Dd	183 ISMELGNPNLSFLKKADIFINGSOLGEDFIQLHKLLRKSTFKNAKLXGPDVGQPRRKTAK 242
Qy	121 MLKSFLKAGGEIVDSVTWHYILNGRTATREDFNLDPDVLDFISSVKVFQVVESTRPKG 180
Dd	243 MLKSFLKAGGEIVDSVTWHYILNGRTATREDFNLDPDVLDFISSVKVFQVVESTRPKG 302
Qy	181 KVMLETSSAYGGAPILLSDTFAAGFMWLDKLGLSARMGLEVNMROVFFFGAGNYHLVDEN 240
Dd	303 KVMLETSSAYGGAPILLSDTFAAGFMWLDKLGLSARMGLEVNMROVFFFGAGNYHLVDEN 362
Qy	241 FDPPLDPYWLSSLFPKCLVGTKVLMASVOGSKRRRLRVYLHCTNTNDNPYKEGDLTLYAINL 300
Dd	363 FDPPLDPYWLSSLFPKCLVGTKVLMASVOGSKRRRLRVYLHCTNTNDNPYKEGDLTLYAINL 422
Qy	301 HNVTKYLRLPYPSFNKOVDKYLLRPLGPHGLLSKSVOLNGITLKQVDDQTLPLMEKPLR 360
Dd	423 HNVTKYLRLPYPSFNKOVDKYLLRPLGPHGLLSKSVOLNGITLKQVDDQTLPLMEKPLR 482
Qy	361 PGSSSLGLPAFSYSFFVIRNAKVAACI 386
Dd	483 PGSSSLGLPAFSYSFFVIRNAKVAACI 508
RESULT 12	
ADZ19006	
ID	ADZ19006 standard; protein; 526 AA.
XX	
AC	ADZ19006;
XX	
DT	16-JUN-2005 (first entry)
XX	
DE	HepHyaluro construct protein.
XX	
Kw	Enzyme engineering; heparanase; hepHyaluro; metastasis;
Kw	autoimmune disease; inflammation; neoplasm; immune disorder;
Kw	antiinflammatory; cytostatic; immunosuppressive; enzyme.
XX	
OS	Synthetic.
XX	
PN	WO2005030962-A1.
XX	
PD	07-APR-2005.
XX	
Pf	17-SEP-2004; 2004WO-EPO10517.
XX	
Pr	26-SEP-2003; 2003US-0506479P.
XX	
PR	20-JAN-2004; 2004US-0537729P.
XX	
PA	(RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX	
PI	Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX	
DR	WPI: 2005-273382/28.
DR	N-PSDB; ADZ19007.
XX	
PT	Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT	nucleotide sequence that encodes mammalian heparanase protein and has two
PT	consensus cleavage sites located between specific nucleotide encoding
PT	residues.

XX Example 2; SEQ ID NO 27; 65pp; English.
PS
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hephyaluro construct
CC protein used in the scope of the invention.
XX

SQ Sequence 526 AA;

Query Match 100.0%; Score 2020; DB 9; Length 526;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
DB 141 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 200
QY 61 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 201 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 260
QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
DB 261 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 320
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 321 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 380
QY 241 FDPDPYWLSSLFPKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 300
DB 381 FDPDPYWLSSLFPKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 440
QY 301 HNVTKYLRPLYPFNSKQVDKYLRLPLGPHGLLSKSVQNLGLTKWVDDQTLPLMEKPLR 360
DB 441 HNVTKYLRPLYPFNSKQVDKYLRLPLGPHGLLSKSVQNLGLTKWVDDQTLPLMEKPLR 500
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 501 PGSSGLGPAFSYFFVIRNAKVAACI 526

RESULT 13
ABB07815
ID ABB07815 standard; protein; 527 AA.

XX ABB07815;

XX 03-JUL-2002 (first entry)

XX Chicken signal peptide/human heparanase chimeric protein sequence.

XX Heparanase; catalytic; cytosolic; antitumor; antibacterial; enzyme;
KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX Synthetic.

XX Gallus gallus.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19

FT /note= "chicken heparanase signal peptide"

FT Protein 20..527

FT /note= "human heparanase mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.
PF
XX 20-SEP-2000; 2000US-00666390.
PR (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;
XX WPI; 2002-338926/37.
DR N-PSDB; ABL40753.
DR
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful
PT to treat various heparin-related disorders and the signal peptide is
PT useful in production of membrane-targeted or secreted recombinant
PT proteins.
XX
PS Disclosure; Page 26-28; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,
CC encoding a polypeptide with heparanase catalytic activity. The signal
CC peptide of the nucleic acid can be used to express membrane-associated or
CC secreted proteins in heterologous expression systems. The encoded
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and
CC invasion, and to intervene with pathologies associated with impaired
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoan and bacterial infections or
CC disintegration of neurodegenerative plaques. The present sequence
CC represents a chicken signal peptide/human heparanase chimeric protein
CC sequence
XX

SQ Sequence 527 AA;

Query Match 100.0%; Score 2020; DB 5; Length 527;
Best Local Similarity 100.0%; Pred. No. 8e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
DB 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 201
QY 61 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 202 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 261
QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
DB 262 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 321
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 322 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 381
QY 241 FDPDPYWLSSLFPKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 300
DB 382 FDPDPYWLSSLFPKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 441
QY 301 HNVTKYLRPLYPFNSKQVDKYLRLPLGPHGLLSKSVQNLGLTKWVDDQTLPLMEKPLR 360
DB 442 HNVTKYLRPLYPFNSKQVDKYLRLPLGPHGLLSKSVQNLGLTKWVDDQTLPLMEKPLR 501
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 502 PGSSGLGPAFSYFFVIRNAKVAACI 527

RESULT 14
ABB02018

ID ABB02018 standard; protein; 527 AA.

XX ABB02018;

XX

DT 12-FEB-2004 (first entry)
 XX Chimeric human-chicken heparanase protein.
 XX
 KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;
 KW enzyme; human.
 XX
 OS Chimeric - Gallus gallus.
 OS Chimeric - Homo sapiens.
 XX
 PN US2003180788-A1.
 XX
 PD 25-SEP-2003.
 XX
 XX 08-MAY-2003; 2003US-00431438.
 PF
 XX 20-SEP-2000; 2000US-00666390.
 PR
 PR 16-AUG-2001; 2001US-00930218.
 XX
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 XX
 XX Goldshmidt O, Pecker I, Vlodosky I, Michal I, Zcharia E;
 PI WPI; 2003-843931/78.
 XX N-PSDB; AAD63532.
 DR
 DR Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful
 PT for treating cancers, microbial infections and aiding wound healing.
 PT
 XX Example; Page 26-28; Opp; English.
 PS
 XX The present invention relates to novel jungle red fowl heparanase protein
 CC and polynucleotides encoding such proteins. Heparanase sequences can be
 CC used to develop treatments for various diseases, to develop diagnostic
 CC assays for these diseases and to provide new tools for basic and directed
 CC research especially in the fields of medicine and biology. They can be
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and
 CC some bacterial and protozoa infections. Recombinant heparanase can also
 CC be used to neutralise plasma heparin, as a potential replacement of
 CC protamine. Sequences of the invention are used in protein therapy. The
 CC present sequence is chimeric human-chicken heparanase protein
 XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 2020; DB 7; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8e-202;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
 DB 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 201
 QY 61 ISWELGNEPNSFLKADIFINGSQGLDIFGLNALLRKSTFKNAKLYGPDVGQPRRTAK 120
 DB 202 ISWELGNEPNSFLKADIFINGSQGLDIFGLNALLRKSTFKNAKLYGPDVGQPRRTAK 261
 QY 121 MLKSFKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
 DB 262 MLKSFKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 321
 QY 181 KVLGETSSAYGGGAPLLSDTFAAGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
 DB 322 KVLGETSSAYGGGAPLLSDTFAAGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 381

QY 241 FDPLPDYWLSSLFPKLVGTQVLMASVQSGSKRRKLRVYLHCTNTDNPYKEGDLTYAINL 300
 DB 382 FDPLPDYWLSSLFPKLVGTQVLMASVQSGSKRRKLRVYLHCTNTDNPYKEGDLTYAINL 441
 QY 301 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQVNGLTLKMWDDQTLPLMEKPLR 360
 DB 442 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQVNGLTLKMWDDQTLPLMEKPLR 501
 QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386
 DB 502 PGSSGLGLPAFSYSFFVIRNAKVAACI 527
 RESULT 15
 ADZ19004
 ID ADZ19004 standard; protein; 527 AA.
 XX
 AC ADZ19004;
 XX
 DT 16-JUN-2005 (first entry)
 XX
 DE HepGS4 construct protein.
 XX
 KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
 KW immunosuppressive; enzyme.
 XX
 OS Synthetic.
 XX
 PN WO2005030962-A1.
 XX
 PD 07-APR-2005.
 XX
 PF 17-SEP-2004; 2004WO-EP010517.
 XX
 PR 26-SEP-2003; 2003US-0506479P.
 PR 20-JAN-2004; 2004US-0537729P.
 XX
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
 PA Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
 PI WPI; 2005-273382/28.
 XX N-PSDB; ADZ19002.
 DR
 DR Synthetic nucleic acid for e.g. inhibitor screening, comprises a
 PT nucleotide sequence that encodes mammalian heparanase protein and has two
 PT consensus cleavage sites located between specific nucleotide encoding
 PT residues.
 XX
 XX Example 2; SEQ ID NO 25; 65pp; English.
 PS
 CC The invention relates to a synthetic nucleic acid molecule that encodes
 CC mammalian heparanase protein, where the nucleic acid comprises two
 CC consensus cleavage sites recognized by endoproteinase. The sequences are
 CC useful for expressing mammalian heparanase in non-mammalian cells and in
 CC inhibitor screening assays for the development of therapeutics or
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
 CC and/or inflammation. This sequence represents a hepGS4 construct protein
 CC used in the scope of the invention.
 XX
 SQ Sequence 527 AA;
 Query Match 100.0%; Score 2020; DB 9; Length 527;
 Best Local Similarity 100.0%; Pred. No. 8e-202;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
 DB 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 201
 QY 61 ISWELGNEPNSFLKADIFINGSQGLDIFGLNALLRKSTFKNAKLYGPDVGQPRRTAK 120

Db	202	ISWELGNPNSEFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGGQPRRTAK	261
Qy	121	MLKSFLKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVQVWESTRPGK	180
Db	262	MLKSFLKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVQVWESTRPGK	321
Qy	181	KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNTHLVNEN	240
Db	322	KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNTHLVNEN	381
Qy	241	FDPDPVWLSLLFKKLVGTVMASVOGSKRRKLRVYLHCTNTDNPYKEGDLTYAINL	300
Db	382	FDPDPVWLSLLFKKLVGTVMASVOGSKRRKLRVYLHCTNTDNPYKEGDLTYAINL	441
Qy	301	HNVTKYLRLPYPFSPNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR	360
Db	442	HNVTKYLRLPYPFSPNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMVDDQTLPLMEKPLR	501
Qy	361	PGSSGLPAPFSYSFFVIRNAKVAACI	386
Db	502	PGSSGLPAPFSYSFFVIRNAKVAACI	527

Search completed: June 5, 2006, 12:09:41
Job time : 78.9552 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model
Run on: June 5, 2006, 12:10:07 ; Search time 12.8514 Seconds
(without alignments)
2889.939 Million cell updates/sec

Title: US-10-645-659A-1
Perfect score: 2020
Sequence: 1 KFKKSTYSRSSVDVLYTFA.....LPAFSYSPFVIRNAKVAACI 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	789.5	39.1	480	2 JC7506	heparanase protein
2	407	20.1	521	2 T45608	hypothetical prote
3	169.5	8.4	190	2 T01953	hypothetical prote
4	105.5	5.2	411	2 S74760	hypothetical prote
5	104.5	5.2	788	1 S00652	phosphoribosylamin
6	103.5	5.1	500	2 D87541	beta-xylosidase [i
7	101.5	5.0	670	2 T38446	microtubule-associ
8	100	5.0	654	2 T14202	NADH2 dehydrogenas
9	98	4.9	644	2 A97268	methionyl-tRNA syn
10	97	4.8	379	2 A69974	cystathionine gamm
11	96	4.8	557	1 ODNC1	cytochrome-c oxida
12	95	4.7	385	2 S28360	Al1 protein - beet
13	95	4.7	511	2 S61166	probable membrane
14	95	4.7	699	2 F95146	DNA topoisomerase
15	95	4.7	701	2 D98014	DNA topoisomerase
16	95	4.7	817	2 H75035	probable membrane
17	94.5	4.7	596	2 T04506	hypothetical prote
18	94.5	4.7	804	2 G71546	probable DNA gyras
19	94	4.7	805	2 C86525	DNA gyrase subunit
20	94	4.7	805	2 H72098	DNA gyrase, chain
21	93.5	4.6	356	2 F64383	hypothetical prote
22	93.5	4.6	883	2 AE0207	conserved hypoteth
23	93	4.6	454	2 T20829	probable serine ca
24	93	4.6	492	2 T3859	uroporphyrinogen I
25	93	4.6	816	2 G71127	hypothetical prote
26	93	4.6	837	1 A31842	endo-1,4-beta-xyla
27	92	4.6	485	2 F64165	hypothetical prote
28	92	4.6	804	2 A81701	DNA gyrase, chain
29	92	4.6	879	2 F81453	DNA-directed DNA p

30	91.5	4.5	649	2 F85682	unknown protein en
31	91.5	4.5	782	2 H90823	probable secreted
32	91.5	4.5	1500	1 JQ1348	carbamoyl-phosphat
33	91	4.5	796	2 D97065	transketolase [imp
34	91	4.5	822	2 F83016	penicillin-binding
35	91	4.5	2013	2 A11489	probable peptidogl
36	90.5	4.5	418	2 H97437	cyclopropane-fatty
37	90.5	4.5	418	2 AC2656	cyclopropane-fatty
38	90.5	4.5	455	2 H84955	UDP-N-acetylmuram
39	90.5	4.5	500	2 E83882	alpha-L-arabinofur
40	90	4.5	604	2 E75119	hypothetical prote
41	90	4.5	847	2 AG1001	nitrite reductase
42	89.5	4.4	419	1 S75867	phosphoribosylamin
43	89.5	4.4	425	2 C97354	hypothetical prote
44	89.5	4.4	828	2 S56250	probable membrane
45	89	4.4	541	1 ODZJ1	cytochrome-c oxida

ALIGNMENTS

RESULT 1

JC7506
heparanase protein 2a - human
C;Species: Homo sapiens (man)
C;Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 09-Jul-2004
C;Accession: JC7506
R;McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircocock, M.; Pat
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000
A;Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me
A;Reference number: JC7506
A;Accession: JC7506
A;Molecule type: mRNA
A;Residues: 1-480 <MCK>
A;Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885
C;Comment: This protein, an intracellular membrane-bound enzyme, has biological and therai
therapies.
C;Genetics:
A;Gene: hpa2a
A;Map position: 10q23-10q24
C;Keywords: heparin binding; membrane bound

Query Match 39.1%; Score 789.5; DB 2; Length 480;
Best Local Similarity 44.0%; Pred. No. 1e-56;
Matches 164; Conservative 61; Mismatches 129; Indels 19; Gaps 5;

QY	30	GLN-ALLRTADLQWSSNAQLLDYCSKG-----YNISWELGNEPNSFLKKADIFING	82
DB	107	GLSPAFURFGGKRTDFLQFNLRNPAKSRGGPGDPDYLLKNYE--DEPNRYRMHGRAVNG	164
QY	83	SQLGEDPIQLHLRK-STFKNAKLYGPDVGQPRRKTKAMLKSLFKAGGEVIDSVTWHRY	141
DB	165	SQLGKDYIQLKSLLOPIRIYSRASLYGFNIGRPKNVTALLDGFMKVAGSTVDATVQHC	224
QY	142	YLNGRTATREDFLNPVDLIDIFISSVQVKVQVVESTRPKKVKWLGETSSAYGGAPLLSOT	201
DB	225	YIDGRVVKWDFLKTRELLDTLSQIRKIQKVNTVTPGKIKWLEGVVTTSAGGTNNLSDS	284
QY	202	PAAGFMWLDKLGLSARMGIEVWMEQVFFGAGNYHLVDENFDPDPYWLSSLPEKLVGTVK	261
DB	285	YAAAGFLWNLNTGLMLANQIDVIRHSFDFHGYNHLVDQNFNPLPDYWLSSLYKRLIGPKV	344
QY	262	LMASVQGSKKR-----KLRYVLHCTNTDNPYKEGDLTVYAINLHNVTKYLRLPYP	312
DB	345	LAVHVAGLQKRPGRVIRDKLRIYACTNNHNNHYVSGSITLFINLHRSKKIKLAGT	404
QY	313	FSNKQVDYLLRPLPGPHGLLSKSVQLNGLTLKMWDDQTLPLPMEKPLPGSGSLGPAFSY	372
DB	405	LRDKLVHQYLLQPYQGEGLSKSVQLNGQPLVMVDDGTLPKLPKRPPLRAGRTLVIPPVTM	464
QY	373	SFFVIRNAKVAAC	385
DB	465	GFFVVKVNNALAC	477

RESULT 2

T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UPI00000A497C; EMBL:ALJ33421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A:Note: F13G24.30

Query Match 20.1%; Score 407; DB 2; Length 521;
Best Local Similarity 32.3%; Pred. No. 2.4e-25;
Matches 137; Conservative 57; Mismatches 158; Indels 72; Gaps 19;
QY 14 DVLTYFANGSGLDLIFGLNALLRTADLQ-----WNSNNAQLLDYCCSKGYNI-SWEL 65
DB 117 DEINSEFLTATGAVVTFGLNALRGHKLKGKAWGAWDHINTQDFLNTVSKGVIDSWEF 176
QY 66 GNEPNSFLKADIFINGSQLGEDFIQLHKLKRSTFKNAKLYGPDVQGP-----RRRTAK 120
DB 177 GNELSG--SGVGASVSAELYKDLIVLKDVINK-VYKNSWLHKPILVAPGFGVEQQWYTK 233
QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRT--ATREDFLNPVDLDFISVQKVP-----QVVE 174
DB 234 LLEI---SGPSVVDVVTHHYLYNGSGNDPALVKKINDPS-----YLSQVSTFKDVAQTQI 286
QY 175 STRPGKVLGWTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMQVFFGAGNY 234
DB 287 EHGPNASPMWGGGAYNSGGRHVSDTFIDSFYLDOLGMSARHNTKVCYCRQLVNG-GFY 345
QY 235 HLVDL-NFDPPLPDYMLSLFKKLVGTVKLVMAVSGSKRRKRLRYLHCTNTDNPYKEGDL 293
DB 346 GLEKGTFTVPNDYYSALLWHRLMGKVLAVQTDGPP--QLRVYAHCSK-----GRAGV 397
QY 294 TLIALNLHNVTKYL-----RUPYPPS-----NKQVDKYLRLP-- 325
DB 398 TLLILNLSNQSDFTVSVNSGINVVLNAESRKKKSLDITLKRPFSGWIGSKASDGLNREEY 457
QY 326 -LGPHG--LLSKSVOLNGLTLKMWDDOTLPLMEKPLRP-GSSLGLPAPSYSPFVIRNAK 381
DB 458 HLTPENGVLRSTKMTVLNGSKLPTATGDIPLSL-EPVLRSVNSPLANPLUSMSFVLNPF 516
QY 382 VAAC 385
DB 517 ASAC 520

RESULT 3

T01953
hypothetical protein T2L5.6 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C:Accession: T01953
R:Geisell, C.; Smith, A.; Le, T.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of A. thaliana T2L5.
A:Reference number: Z14470
A:Accession: T01953
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190 <GEI>

A:Cross-references: UNIPROT:O82604; UNIPARC:UPI00000A8F7D; EMBL:AF096371; NID:G3695386; I
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 36/2; 69/3
A:Note: T2L5.6
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 8.4%; Score 169.5; DB 2; Length 190;
Best Local Similarity 27.8%; Pred. No. 1.6e-06;
Matches 54; Conservative 34; Mismatches 57; Indels 49; Gaps 9;
QY 225 QVFFGAGNYHLVD-ENFDLPDYLWLSLLFKKLVGTGTVKLVMAVSGSKRRKRLRYLHCTNT 283
DB 12 QSLIG-GNYGLLNTFTNPDPYYSALIWRLMGRKALFTTSGTK--KIRSYTHCA-- 66
QY 284 DNPYKEGDLTYAINLHNV-----TKYLRLPYPPFSNKQVDKYLRLPL 326
DB 67 ---RQSKG-ITVLLMLNDNTTVVAKVELNNSFSLRHTKHKM-----SYKRASSQLFG-- 115
QY 327 GPHGLL-----SKSVOLNGLTLKMWDDOTLPLMEKPLRPGSSGLGLPAFS 371
DB 116 GPNQVIOREYHILTAQDCNLHLSQTMLLNGALQVNSMGDLPIPIPIHINSTEPIIAPYS 175
QY 372 YSFVIRNAKVAAC 385
DB 176 IVFVHMENVVPAC 189

RESULT 4

S74760
hypothetical protein slr1617 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S74760
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
Okumura, K.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
O. Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74760
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <KAN>
A:Cross-references: UNIPROT:P72895; UNIPARC:UPI00000C0C3B; EMBL:D90901; GB:AB001339; NID:
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 5.2%; Score 105.5; DB 2; Length 411;
Best Local Similarity 21.4%; Pred. No. 0.85;
Matches 66; Conservative 54; Mismatches 95; Indels 93; Gaps 17;
QY 2 KFKNSTYSRSSVDVLYTFANCGLDLIF---GLNALLRTADLQWNSNAQL----- 49
DB 151 EFLISPTREQIDI---FAGSTKLDLASEENIDCVHLANPRVYTSNVAMGQTLTLRN 207
QY 50 LLDYCSSKG---YNISWEL-----GNPNPSFLKK-----ADIFINGSQJCE 87
DB 208 VIDVCLAKDIPLYPSSWEIYSGYAGTIHADSTPALPRGPYGETKYLAELI----- 260
QY 88 DFIQLHLKLRKSTFKNAKLYGPDVQGPERRKTAKMLKSLKAGGEVIDSVTHHYLYNGRT 147
DB 261 DHCRRRTGLKAILRISPSVYSGMSDKP-----KFIFFFKASQOQKIVT--HHYING-- 311
QY 148 ATREDFLNPDV---LDIFISSVQKVFQVVESTRPGKKVILGETSSAYGGGAPLLSDTFA 203
DB 312 -----NPKLDDLHIDDLISSIVATL-----KSRFIGNLNI-----GTGQLSSTLK 351
QY 204 AGFWMLDKGLSA-----RMGIEVVMQVFPFGAGNYHLVDENFDLPDYLWLSLLFKKLVG 258
DB 352 IAEIMRDELGGSSMIQOIEVNTVEVATMNYGRAN-HVLD-----WEPVIFFE-QG 400

D87541
beta-xylosidase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C:Accession: D87541
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.;
N.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon-
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: UNIPROT:Q9A5U0; UNIPARC:UPI00000C76D1; GB:AE005673; NID:g13423886; PJ2
C:Genetics:
A:Gene: CC2357

Query Match 5.1%; Score 103.5; DB 2; Length 500;
Best Local Similarity 25.2%; Pred. No. 1.6;
Matches 55; Conservative 36; Mismatches 80; Indels 47; Gaps 15;

QY 8 YSRSSVDVLYTFANCSGLDLIFGLN---ALLRTAD---LOW--NSSNAQL-----LLD-- 52
DB 81 YDMTKIDQLYDALLAKGIKPFIELGFTPEANKTSDQTIFYWKGNTSHPKLGPWRDLIDAF 140
QY 53 -YCSSKGYNI-----SW--ELGNEPN--SFLKKADIFINGSQLGEDFIOLHKLKRSTFKN 103
DB 141 VHLRLARYGVEEVRTWTFEEVWNEPNLGFWEKAD-----QAAYFELYDV---TARA 188
QY 104 AKLYGPD--VGQPRRKTKAKLKSFL---KAGEVIDSVTHHHYLLNG-----RTATREDFL 154
DB 189 IKAIIDPSLRVGGPATAGAAWPEFLAHVKKSGSAVDFTTHTYGVGGFLDEKGVQDTKL 248
QY 155 NPDVLDIFISSQVKVQFQVVE--STRPGKKVWLGETSSAY 191
DB 249 SPSP-DAVGVGVRRVREQIEASAFPLGLPLFYFTWSTSY 285

RESULT 7
T38446
microtubule-associated protein ssm4 - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T38446; T00012
R:McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z21793
A:Accession: T38446
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-670 <MCD>
A:Cross-references: UNIPROT:O42667; UNIPARC:UPI0000135FDD; EMBL:AL009227; PIDN:CAA15832.J
A:Experimental source: strain 972H-; cosmid c27D7
R:Yamashita, A.; Watanabe, Y.; Yamamoto, M.
Genes to Cells 2, 155-166, 1997
A:Title: Microtubule-associated coiled-coil protein Ssm4 is involved in the meiotic devel
A:Reference number: Z14042; MUID:97311255; PMID:9167972
A:Accession: T00012
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-670 <YAM>
A:Cross-references: UNIPARC:UPI0000135FDD; EMBL:AB000269; NID:g3341860; PIDN:BAA31857.1;
C:Genetics:
A:Gene: ssm4; SPAC27D7.13c
A:Map position: 1

Query Match 5.0%; Score 101.5; DB 2; Length 670;
Best Local Similarity 23.5%; Pred. No. 3.6;
Matches 72; Conservative 42; Mismatches 109; Indels 83; Gaps 16;

QY 6 STYSRSRSDVLYTFAN-----CSGLDIFGLNALLRTADLQWNSNAQLLDYCS----- 55

Db 148 STEELSSFT--TLNMSDTSKLSGLD-----DSFMEEBEFVWVDN---VLQCEKXFTF 197
QY 56 -SKGYNISWELGNEPNSFLKKA---DIFNGSOLGEDFQLHKLRLK---STFKNAKLY 107
Db 198 HSKGSYLKENLASE-----LRKRLDELMCENTALKKIDKLNKELEKVEPQLTFLRSK-- 251
QY 108 GPDVGQPRR-KTAKMLKSLFKAGGEV-----IDSVTHHHYLLNGRTATREDFL 154
Db 252 -NSIEKPRNFRREKFLKFLAMQKIKYLKRLKQLRQIPNYKYSRSLNSKTPKQSDNW 310
QY 155 NPDVLD---IFTSVQKVFQVVESTRPGKKVWLGETSAYGGAPLLSDTFAAGFWMLDK 211
Db 311 TTQVTPSSLLGVSEVSKVLQL-----KQVQVDITE-----LVKIPK 346
QY 212 LGLSARMGTEVVMQVFFGAGN-----YHLVDENFDPLPDYWLSSLFKKLGVTKVLMAVQ 267
Db 347 NPFSEKLITSNVRYNLNIVPGSLDLOFSLTNENF-----VHWNSTVYQELLNLKSNNSVD 402
QY 268 GSKRRK 273
Db 403 GVKTRR 408
RESULT 8
TL14202
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Rhizopus stolonifer mitochondria
C:Species: mitochondrion Rhizopus stolonifer
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: TL14202
R:Paquin, B.; Roewer, I.; Wang, Z.; Lang, B.
submitted to the EMBL Data Library, November 1994
A:Description: A robust fungal phylogeny using the mitochondrially encoded nad5 protein
A:Reference number: Z17884
A:Accession: TL14202
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-654 <PAQ>
A:Cross-references: UNIPROT:P50367; UNIPARC:UPI0000113081A; EMBL:U17011; NID:g604923; PID
A:Experimental source: DAOM 148428
C:Genetics:
A:Genome: mitochondrion
A:Note: nad5
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 5.0%; Score 100; DB 2; Length 654;
Best Local Similarity 21.5%; Pred. No. 4.6;
Matches 82; Conservative 60; Mismatches 139; Indels 100; Gaps 17;
QY 30 GLNALLRTADLQWSSNAQLLDYCSCSKGYNISWELGNEPNSFLKKADIFNGSOLGEDF 89
Db 256 GYVLLRSPLEFSGTALLITWVGALTAFPAATGGLQND-LKRVIAYSTCSQLGLLF 314
QY 90 -----IQLHKLRLKSTFK-----NAKLYGPDVGQPRRKRTAKMLK-----SFLK 127
Db 315 LVCGLSQYNVALFHLVNHAWFALLELSAGSVIHANDEQDLRKFGGLRLLPFTYSMMV 374
QY 128 AGG-----EVIDSVTHHHYLLNGRTATREDFLNPDVLDIFISSVQKVFQV 172
Db 375 IGSLSLMALPPLTGTGFYSKDLIITELAYGHVSFSGN-----LVYWLASVAAVFTA 422
QY 173 VESTRPGKKVWLGETSAYGGG-----APLSS-----DTFAAGFWMLDKLGLS 215
Db 423 MYSIRSLVLTFLG-----YPNGPKINYNINHEAPLIMAIPLVVLAVERISFFGVYTK-DLF 476
QY 216 ARMGTEVVMQVFFGAGNTHLVNFDPLPDYWLSSLFKKLGVTKVLMAVQSKRRKL 275
Db 477 VGMGDTFYNNALFIHPNHSILVDTEFG-LP---MSMKFLPLGLSLGTGTV-----LA 525
QY 276 VY-----LHCTNTDNPRYKEGDLTYAINLHN--VTKYLRLPYPFSNKOVDKYLLR 324
Db 526 IYWIFDELPNKFISTKLGRIYRFFNQKYFYFDNIYNNLLNKLFLNFGYT-TNKILDRGATE 584

QY 325 PLQPHGLL-----SKSVQVNLGL 341
Db 585 LVGPYGLVNVFKSASNKVSGL 605

RESULT 9

A97268
methionyl-tRNA synthetase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C:Accession: A97268

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: A97268

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 <KUR>

A:Cross-references: UNIPROT:Q97EW5; UNIPARC:UPI000013658D; GB:AE001437; PIDN:RAK80932.1;

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2991

C:Superfamily: methionyl-tRNA synthetase, dimer-forming

Query Match 4.9%; Score 98; DB 2; Length 644;

Best Local Similarity 18.4%; Pred. No. 6.6;

Matches 78; Conservative 57; Mismatches 146; Indels 142; Gaps 19;

QY 6 STYSRSVDVLYTFANGSGLDLIFGLNALLRTADLQWSSNAQLLDYCSCSKGYNIS--- 62

Db 24 NTYTVASDALVRFKRLQGYDAFW-----LTGDEHGQKIQRIAEKDGITPKAYV 73

QY 63 -----WELGN-EPNSFLKKADIFNGSOLGEDFQLHKLRLKSTFKNAKLYG--- 108

Db 74 DEIVAGIKDLWKMMNITSYDKFIRTTD-----EBHVRVAVQKIVKFKFYDNGDIYSAY 124

QY 109 -----PDVGQPRRKRTAKMLKSF--LKAGGEVIDSVTHHHYLLN 144

Db 125 EGMVYTCESFWTETQLVDGKCPDGPVKEKTEEAYFFKMSKYADRLIKYIEHPDPFIQ 184

QY 145 GRATATRE---DFLNPDPVLDIFISSVQKVFQVVESTRPGKKVWLGETSAYGGAPLLSDT 201

Db 185 PESRKNEMLNFLKPGQLDLCIS-----RSSFDMGIPITFDE 221

QY 202 FAAGFWMLDKL-GLSARMGIEVVMQV--FGAGNYHLVDEN---FDPLPDYWLSSLF-- 253

Db 222 KHVIYVWIDALSNIYITALGYSDNDELYNKFWPADLHVGKDIIRFHTI--YWPIMLMAL 279

QY 254 -----KKLVGTVKVLMAVQSKRRKL------VYLHCTNTDNPRYKEGDLTYAINLH 301

Db 280 DLPLPKQVFGHGWL--VDGGMKSKGNVDPVVLNINEFGTDPVY----- 324

QY 302 NVTKYLRLPYPSFN-----KQVDKYLRPLPGHLLSKSVQVNLGLTLKMWDDQTL 351

Db 325 -----YLLHEIPFGSDGLFNNEIFIKKINSDLANDLG--NLVSRTA---AMIEKFPDGSIQ 375

QY 352 PPL 354

Db 376 PPV 378

RESULT 10

A69974

cystathionine gamma-synthase homolog yrhB - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Oct-2004

C:Accession: A69974

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.

Date: 7/6

Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallei, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, A.; Authors: Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, K.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: A69574
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-379 <KUN>
 A:Cross-references: UNIPROT:O05394; UNIPARC:UPI0000060808; GB:Z99117; GB:AL009126; NID:9
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yzhB
 C:Superfamily: cystathionine gamma-synthase

Query Match 4.8%; Score 97; DB 2; Length 379;
 Best Local Similarity 21.7%; Pred. No. 3.7;
 Matches 72; Conservative 54; Mismatches 112; Indels 94; Gaps 16;
 QY 73 LKQADIFINGSQIGED-----FIQLHKLKSTFKNAKLVGPDVGQPRRTAKMLSKFLK 127
 DB 1 MKKKTLMIHGGITGDEKTAGVSPVQV---STYQPKA-GQHTGYEYSRTNPTRTALE 56
 QY 128 AGEVLDSTVTHHYLYNGRTATREDFLNPDVLDIFSSVQKVQVVESTPRGKKVWLGET 187
 DB 57 ALVTELESGEAGYAFSGMAA-----ITAVMLFP-----NSGDHVVIL--T 94
 QY 188 SSAYGSGAPLLSDTEAAGFMWLDKGL-----SARMGLEVMVR---QVFFGAGNVHLV 237
 DB 95 DVIYGG-----TYRMTKVLNRLGIESTFVTSREVEKAIKRNPKAIY-----I 140
 QY 238 DENFDLPDYLWLSLLFKLVGTVKLVMSVQSKRRKLRVYLHCTNDNPRYKGDITLYA 297
 DB 141 EPTNPL-----LKITDLTMADIAKAGVLLIVDNTFNTFYFQOPLTLGA 186
 QY 298 -INLHNVTYKLRPLPYF-----SNKQVDKYL-----RPLGPHG--LLSKSVQLNG 340
 DB 187 DIVLSATKYLGSHSDVGVGLVVTASKELGEELHFQVNSTGGVGLGPDQSDWLLMRGIKTLG 246
 QY 341 LTLKWDDQTLF-----PLMEKPLRPGSS 364
 DB 247 LRMEAIDQNARKIASFLENHPAVQTLTYFGSS 278

RESULT 11
 ONC1
 cytochrome-c oxidase (EC 1.9.3.1) chain I - *Neurospora crassa* mitochondrion
 N:Alternate names: cytochrome a3 polypeptide I; cytochrome a3 polypeptide I
 C:Species: mitochondrion *Neurospora crassa*
 C:Date: 13-Aug-1986 #sequence revision 13-Aug-1986 #text_change 09-Jul-2004
 C:Accession: A04469; S07650; S07651
 R:Burgen, G.; Scriven, C.; Machleidt, W.; Werner, S.
 EMBO J. 1, 1385-1391, 1982
 A:Title: Subunit 1 of cytochrome oxidase from *Neurospora crassa*: nucleotide sequence of
 A:Reference number: A00469; MUID:84207889; PMID:6327266
 A:Accession: A00469
 A:Molecule type: DNA
 A:Residues: 1-557 <BUR>
 A:Cross-references: UNIPROT:P03945; UNIPARC:UPI0000127F8D; EMBL:X01850; NID:g13119; PIDN
 A:Note: the amino end of the mature protein may be 3-Ser
 R:Field, D.J.; Sommerfeld, A.; Saville, B.J.; Collins, R.A.
 Nucleic Acids Res. 17, 9087-9099, 1989
 A:Title: A group II intron in the *Neurospora* mitochondrial *coi* gene: nucleotide sequence
 A:Reference number: S07649; MUID:90067912; PMID:2531370
 A:Accession: S07650

A:Molecule type: DNA
 A:Residues: 1-73 <PIE>
 A:Cross-references: UNIPARC:UPI0000017215B; EMBL:X14669; NID:g13123
 R:Collins, R.A.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S07651
 A:Accession: S07651
 A:Molecule type: DNA
 A:Residues: 1-526, V, 528-540 <COL>
 A:Cross-references: UNIPARC:UPI0000017215C; EMBL:X14669; NID:g13123
 R:Vassiliev, A.O.; Plesefsky-Vig, N.; Brambl, R.
 Proc. Natl. Acad. Sci. U.S.A. 92, 8680-8684, 1995
 A:Title: Cytochrome c oxidase in *Neurospora crassa* contains myristic acid covalently linked
 A:Reference number: A43101; MUID:96004602; PMID:7567996
 A:Contents: annotation; modified site
 C:Genetics:
 A:Gene: COI
 A:Genome: mitochondrion
 A:Genetic code: SGC3
 A:Introns: 72/1; 99/2; 210/3; 249/2
 C:Function:

A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecules
 ns from the mitochondrial matrix producing two molecules of water and lowering the concn
 A:Pathway: oxidative phosphorylation; respiratory chain
 A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-mem
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium;
 ducase; respiratory chain; transmembrane protein
 F:16-462/Domain: cytochrome-c oxidase chain I homology <COI>
 F:67,383/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:246,295,296/Binding site: copper (His) #status predicted
 F:246-295/296/Binding site: copper (His) #status predicted
 F:250/Binding site: oxygen (Tyr) #status predicted
 F:324/Binding site: myristate (Iys) (covalent) #status experimental
 F:373/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:381/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 4.8%; Score 96; DB 1; Length 557;
 Best Local Similarity 20.9%; Pred. No. 7.8;
 Matches 68; Conservative 43; Mismatches 116; Indels 98; Gaps 16;
 QY 99 STFNKALYGPVQGPQRRTAKMLSKFLKAGEVIDSVTW-HHYLYNGRTATREDFLNPD 157
 DB 265 SAYSNKSVFG-YIG-----MYAMMSIG--ILGFIVSHHMYTVGLDVTDFYTTAA 313
 QY 158 VLDIFISGVQVQVVESTPRGKKVWLGETSAYGGGAPLL-SDTFAAGFMWLDKLG--- 213
 DB 314 TLIIAVPTGIKIFS-----WL---ATCYGGSRLTPSMLFALGFVFMFTIGGLS 359
 QY 214 --LSARMGLEVMVRQVFF-----GA-----GNYHLVDE---NFDPPL---PDY 247
 DB 360 GVVLNANSLDIAFHDTYVVAHFHVLMSGAVFAMFSGWYHVPKILGLNYNVLKSAQF 419
 QY 248 WLSLLFKLVGTVKLVMSVQSKRRKLRVYLHCTNDNPRYKGE----- 291
 DB 420 WLLFVGLNLTFFPQHFGLGQMPRR-----ISDYPDAFSGWNLSSFGSIVSVVAS 470
 QY 292 DLTLYANLHNVTYKLRPLPYFPFSKNQVDKYLRLPLGPHGLLSKSVQLNGLTLKWVDDQTL 351
 DB 471 WLFYIVYIQLVQGEYAGRPWSIPQFYTDSL-----ALLNRSY-----PSL 513
 QY 352 PPLMEKPLRPGSSLSGLPAFYSYFV 376
 DB 514 EWSISSPPKPHSFASLPLOQSSFFL 538

RESULT 12
 S28360
 A11 protein - beet curly top virus
 C:Species: beet curly top virus
 C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004
 C:Accession: S28360
 R:Stanley, J.; Markham, P.G.; Callis, R.J.; Pinner, M.S.

EMBO J. 5, 1761-1767, 1986
A:Title: The nucleotide sequence of an infectious clone of the geminivirus beet curly to
A:Reference number: S28360
A:Accession: S28360
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-385 <STA>
A:Cross-references: UNIPROT:Q39485; UNIPARC:UPI00000F979C; GB:M24597; EMBL:X04144; NID:9
C:Superfamily: tomato golden mosaic virus AL1 protein

Query Match 4.7%; Score 95; DB 2; Length 385;
Best Local Similarity 21.8%; Pred. No. 5.6;
Matches 67; Conservative 43; Mismatches 98; Indels 100; Gaps 18;
Qy 92 LHKLLR---KSFKNKLYGPDVGP-----RRKTAKMLKSLFKAGGEVDSVT 137
Db LHVLLQLSEKVOITNIRLF-DLVSPTRSAHPNIRQAKSSDVKYVDKG---DTTE 139
Qy 138 WHYYLNGRTA-----TRED-----FLNPDVLD-----IFISSVQKV 169
Db WGEFQIDGRSARGGQQTANDSYAKALNATSLQALQILKEEQKDYFLQHNLNNAQKI 199
Qy 170 FQVVESTPGKKW-----LGETSSAYGG-----GAPLL-----SDTFA 203
Db 200 FQ-----RP-PDPWTFPLFLSSFTNVPEEMQEWADAYFGVDAARPLRYNSIIVEGDSRT 253
Qy 204 AGFWMLDKLG---LSARMGIEVVMQVFFGAGNHYHLVDENPDPLDYWLSSLFKKLVT 259
Db 254 GKTMMARSLGANYITGHLDFS---PRYYDEVENVID---DVPDYLKMKHKGHLIGA 307
Qy 260 -KVLMSVQGSRRKRLRVYLHCTNTDNP-----RYKEGLTLYAINLHNVTYKRL 309
Db 308 QKEWQTNLYKGPVKVIGKIPICILLNCPGESSYQQFLEKPENEALKSWTLHNS-FCKL 366
Qy 310 PYPFSNKQ 317
Db 367 QGPLENNQ 374

RESULT 13
S61166
probable membrane protein YDR371w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D9481.7
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S61166
R:Dirig, H.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of S. cerevisiae cosmid 9481.
A:Reference number: S61159
A:Accession: S61166
A:Molecule type: DNA
A:Residues: 1-511 <DIN>
A:Cross-references: UNIPROT:Q06350; UNIPARC:UPI000006A221; EMBL:U28373; NID:9849184; PID
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Gene: MIPS:YDR371w
A:Cross-references: SGD:S0002779
A:Map position: 4R
C:Superfamily: Serratia marcescens chitinase
C:Keywords: transmembrane protein
F:18-34/Domain: transmembrane #status predicted <TMM>

Query Match 4.7%; Score 95; DB 2; Length 511;
Best Local Similarity 19.0%; Pred. No. 8.4;
Matches 75; Conservative 70; Mismatches 149; Indels 100; Gaps 17;
Qy 23 SGLDLIFGLNALLRTRADLQWSSNAQL-----LLDYCSSKGYNLSWELGNE 68
Db 124 SLENNLYKSLAIKNSLTKSSNNVQNILPLGCTIGELFYLNKTCSDKKFKVIMSIGW 183
Qy 69 PMSFLKKADIFNGSLQGDFFIQ-----LHKLLRKSTFKNKLKLYGPDVGQPR--RKTAQM 121

Db 184 SDS--ENFKIILKDDKLLQLQNFVDSVETMPLRGLPDGIDLDWEFFGNNESEPRGYLKLRM 241
Qy 122 LKGFLLKA-GGEVIDSVTHHYLYNGRTRATREDFLNPDVLDIFISSVQKVQFQVWVSTRPGK 180
Db 242 LRLKNSLESQIFGKRTEDHFQLSIAAPAFK-----KLFYLPITFIDQYVDYNNMT 294
Qy 181 KWLGETSSAYGGGAPLLST-----PAAGFMWLDKLGSLARMGIEVVMQVFFGAGNYH 235
Db 295 YDYGSWSETTYGHSNLSFSELTNGNFAMHYM-IDRFQVNSR---KVLGMAAYGR-SFH 349
Qy 236 LVDENFDPLD--YWLSLLEFKL-----VGT-----KVLN 263
Db 350 IKDNKEFFNQNTLVNLIKFGVKGPKEIDKADGEGIMPYKNLPKIGTIEQIDPKYVS 409
Qy 264 ASVOGSKRRKRLRVYLHCTNTDNPYKEGDLT-----LYA-----INLHN 302
Db 410 AYCDEKN---SIFISVDNTKSVTKAEYTHNNLGGFWWESCGEAYANERSLINA 466
Qy 303 VTKYLR-----PYPFSNKQVDKYLRLPLGPHGLLS 333
Db 467 EGLHFNVSXKPSIFQDVRVKYLYNKYDGGFLS 500

RESULT 14
F95146
DNA topoisomerase I [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: F95146
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-699 <KUR>
A:Cross-references: UNIPROT:Q97QF3; UNIPARC:UPI000003BFDE; GB:AE005672; PIDN:AAK75367.1;
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPl263
C:Superfamily: DNA topoisomerase I

Query Match 4.7%; Score 95; DB 2; Length 699;
Best Local Similarity 24.0%; Pred. No. 13;
Matches 60; Conservative 39; Mismatches 75; Indels 76; Gaps 17;
Qy 103 NAKLYGPDVGQPRRTAKMLKSLKAG---GEVIDSVTH-HYVYN----- 144
Db 67 NIRKGGLINDLKKEAKKANKVFLASPDREGE---AISWHLAHLNLDENDANRVFNE 123
Qy 145 -GRATATREDFLNPDVLDIFISSVQKVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFA 203
Db 124 ITKDAVKNAFKPRKIDMDLVDAQARRILDR-----VGYSIS-----PIL----- 165
Qy 204 AGFWMLD-KLGLSARMGIEVVMQVFFGAGNHYLVD-EN-----FDPLPDYW-LSLLFKKL 256
Db 166 ----WKKVKKGLSAGRVQSIALKLI-----IDRENEINAFQ-EEYWTVDVAFKK- 210
Qy 257 VGTKVLMSVQGSRRKRLRVYLHCTNTDNPYKE- -GDLTLYNLHNVTYKLR-----L 309
Db 211 -GTKQFHASFYGVDRKQMKL-----TSNNEVKVELSLRTSKDFSVQVDKKERKNAPL 263
Qy 310 PYPFSNKQVD 319
Db 264 PYTSSNQMD 273

RESULT 15
D98014

